

**ФЕДЕРАЛЬНОЕ АГЕНТСТВО ВОЗДУШНОГО ТРАНСПОРТА
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ГРАЖДАНСКОЙ АВИАЦИИ»**

**Кафедра аэродинамики, конструкции и прочности летательных ап-
парат**

М.С. КУБЛАНОВ

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

**МЕТОДОЛОГИЯ И МЕТОДЫ РАЗРАБОТКИ
МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ
МЕХАНИЧЕСКИХ СИСТЕМ И ПРОЦЕССОВ**

Часть II

Планирование экспериментов и обработка результатов измерений

Издание третье,
переработанное и дополненное

Рекомендовано УМО вузов РФ
по образованию в области экс-
плуатации авиационной
и космической техники
в качестве учебного пособия

МОСКВА 2004

УДК 519.876.5(075.8)
ББК 22.2в631.0я73-1
К88

Печатается по решению редакционно-издательского совета
Московского государственного технического университета ГА

Рецензенты: д-р техн. наук, проф. В.Г Ципенко;
канд. техн. наук, проф. С.Г. Косачевский (проректор по на-
учной работе Ульяновского высшего авиационного училища
ГА)

Кубланов М.С.
К88 Математическое моделирование. Методология и методы разработки ма-
тематических моделей механических систем и процессов. Часть II. Пла-
нирование экспериментов и обработка результатов измерений. Издание
третье, переработанное и дополненное: Учебное пособие. – М.: МГТУ
ГА, 2004. – 125 с.: ил. 14, табл. 36.
ISBN 5-86311-437-1

Книга представляет собой учебное пособие, предназначенное для студентов, знакомых с высшей математикой в объеме первых двух курсов вузовского образования. Данная книга является второй частью пособия, в которой излагаются отдельные вопросы математической статистики, необходимые при планировании экспериментов, а также методы обработки и анализа информации, предназначенные для изучения в процессе магистерской подготовки. Большинство сложных тем изложено на подробно разобранных примерах. Особое внимание уделено связи практических задач с объектами прикладной математики.

Данное учебное пособие издается в соответствии с учебными планами для студентов специальностей 130300, 330500 и направления 552000 всех форм обучения.

Рассмотрено и одобрено на заседаниях кафедры АКПЛА 10.02.04 г. и методических советов по специальности 130300 16.02.04 г., по специальности 330500 20.02.04 г., по направлению 552000 17.02.04 г.

К $\frac{1602110000-028}{Ц33(03)-04}$

ББК 22.2в2.2в6к73-1
К88

Св. тем. план 2004 г.

поз. 28

© Московский государственный технический университет ГА, 2004
© Кубланов М.С.

ОГЛАВЛЕНИЕ

Часть II. Планирование экспериментов и обработка результатов измерений	
Предисловие	4
Раздел 2	
МЕТОДЫ ОБРАБОТКИ ИНФОРМАЦИИ	
Глава 5. Основы теории вероятностей и математической статистики	5
5.1. Основные термины теории вероятностей и математической статистики.....	5
5.2. Отбор информации.....	12
5.3. Точечные оценки.....	14
5.4. Законы распределения.....	20
5.5. Интервальные оценки.....	23
5.6. Проверка статистических гипотез.....	25
5.7. Статистическая проверка адекватности математических моделей.....	33
5.8. Основы статистического контроля качества технологических процессов.....	39
Глава 6. Основы многомерного статистического анализа	46
6.1. Классификация задач многомерного статистического анализа.....	46
6.2. Понятие о корреляционном анализе.....	50
6.3. Дисперсионный анализ.....	53
6.4. Регрессионный анализ.....	59
6.5. Понятие о конфлюэнтном анализе.....	67
6.6. Непрерывные случайные величины и понятие о теории фильтрации.....	68
Раздел 3	
ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА	
Глава 7. Статистические методы планирования эксперимента	72
7.1. Проблемы построения эксперимента.....	72
7.2. Назначение плана эксперимента.....	76
7.3. Планирование объема эксперимента.....	79
7.4. Планирование однофакторного эксперимента.....	85
7.5. Планирование двухфакторного эксперимента.....	87
7.6. Планирование многофакторного эксперимента.....	88
7.7. Неполные и неортогональные планы.....	92
7.8. Сравнение приемов планирования эксперимента.....	97
Глава 8. Особые методы планирования эксперимента	107
8.1. Специальные приемы планирования эксперимента.....	107
8.2. Методы экспертных оценок.....	111
Список литературы	117

Предисловие

Данная книга является второй частью учебного пособия, состоящего из двух частей. В ней сохранена сквозная нумерация разделов, глав, параграфов, рисунков и таблиц, продолжающая нумерацию первой части. Первая часть посвящена собственно теории математического моделирования и полностью изучается студентами механического факультета МГТУ ГА в рамках дисциплины "Моделирование систем и процессов". В ней излагаются основные понятия теории, классификация моделей, дается обзор методов разработки математических моделей и вычислительных методов.

Вторая часть посвящена методам прикладной математики, применяемым при математическом моделировании, и изучается студентами в рамках дисциплины "Планирование экспериментов и обработка результатов измерений". Однако отдельные разделы главы 5 второй части пособия необходимы студентам при изучении дисциплины "Моделирование систем и процессов". Студенты магистерской подготовки изучают отдельные разделы данной части пособия, не входящие в программу дисциплины "Планирование экспериментов и обработка результатов измерений".

Раздел 2

МЕТОДЫ ОБРАБОТКИ ИНФОРМАЦИИ

Приступая к изучению реально существующего объекта, мы сталкиваемся с вопросом: какая информация и в каком виде нам необходима. Любой изучаемый объект многообразен. Число факторов, влияющих на него, бесконечно. Проявления этих факторов могут быть неоднозначны или случайны. Как выбрать основное? Общие соображения для ответа на этот вопрос может дать четко сформулированная цель исследований. Такие соображения можно назвать *постановкой задачи отбора информации*

Более конкретное представление о наборе рассматриваемых факторов, об их количественных характеристиках может быть получено только в результате нескольких **последовательных** приближений эксперимента, сопровождающихся постепенным "созреванием" модели объекта. При этом для достижения даже четко сформулированной цели исследований не всегда ясно, когда можно остановиться, а когда необходимо углублять свое представление об изучаемом объекте, с какой точностью и как необходимо регистрировать данные наблюдений. Было бы очень кстати иметь аппарат для **строгого обоснования отбора информации**.

Таким образом, прослеживается аналогия между объектом наблюдения и зарегистрированными результатами, с одной стороны, и понятиями генеральной совокупности и выборки, с другой. Т.е. можно почерпнуть недостающий исследовательский аппарат из математической статистики.

Удобное представление результатов расчетов и экспериментов подразумевает не только компактность, но и информативность, а также возможность получения *достоверных* выводов по результатам анализа, т.е. *адекватных моделей*. Это понятие достоверности опять приводит нас к необходимости применения **статистических методов** для отбора и обработки информации.

Математическая статистика опирается на аппарат теории вероятностей, поэтому очередная глава посвящена основам теории вероятностей в том объеме, который необходим для понимания следующих глав учебного пособия.

Глава 5. Основы теории вероятностей и математической статистики

5.1. Основные термины теории вероятностей и математической статистики

Теория вероятностей – наука, изучающая закономерности в случайных явлениях.

Случайное явление – явление, которое при неоднократном воспроизведении одного и того же опыта протекает по-разному.

Событие – всякий факт, который в результате опыта может произойти или не произойти.

Если некоторое событие \mathcal{A} заведомо не может произойти (например, температура воздуха не может принять значение -300°C), то такое событие называется невозможным.

Если некоторое событие \mathcal{A} обязательно происходит (например, температура воздуха принимает значение в интервале от -300°C до 1000°C), то такое событие называется достоверным.

Два события называются несовместными, если их одновременное (совместное) наступление невозможно (например, невозможно выпадение одновременно 3 и 5 очков на игральной кости).

Несколько событий образуют полную группу событий, если обязательно происходит хотя бы одно из них, т.е. никаких других, неучтенных, событий быть не может (например, полную группу событий составляют выпадение 1, 2, 3, 4, 5, 6 очков на игральной кости).

Из практики мы знаем, что при повторении одного и того же опыта (например, замер температуры воздуха, бросание игральной кости) получаются различные результаты. Это является следствием влияния неучтенных в данном эксперименте факторов. Будем называть результаты этих опытов *исходами* или *элементарными событиями*.

Рассмотрим идеализированную систему исходов, обладающих следующими свойствами:

- число исходов конечно;
- все исходы образуют полную группу событий;
- все исходы попарно несовместны;
- все исходы равновозможны.

Если появление некоторого исхода влечет за собой происхождение события \mathcal{A} , то такие исходы называют *благоприятными* появлению события \mathcal{A} (например, выпадение 2, 4 или 6 очков являются исходами, благоприятными появлению четного числа).

На этой идеализированной схеме можно дать классическое определение вероятности: вероятность $P(\mathcal{A})$ случайного события \mathcal{A} – это **числовая характеристика** возможности этого события, определяемая отношением $\frac{m}{n}$, где n –

число **всех** равновозможных, несовместных исходов, образующих полную группу событий, а m – число исходов, **благоприятных** появлению события \mathcal{A} . Так, например, вероятность появления четного числа очков при бросании игральной кости определится дробью: $P(\mathcal{A}) = \frac{m}{n} = \frac{3}{6} = \frac{1}{2}$.

Легко видеть основные свойства вероятности:

- вероятность принимает значения от 0 до 1: $0 \leq p \leq 1$;
- невозможное событие имеет нулевую вероятность ($p = 0$);
- достоверное событие имеет единичную вероятность ($p = 1$).

Случайной величиной называют величину, которая в результате опыта может принять только **одно из множества** возможных значений, заранее не известно какое.

Различаются дискретные (принимающие отдельные, изолированные перечисляемые значения) и непрерывные (возможные значения заполняют некоторый промежуток числовой оси) *случайные величины*.

Основное прикладное значение в теории вероятностей имеют законы распределения случайных величин, ставящие в соответствие каждому значению случайной величины вероятность именно его появления.

Если случайная величина ξ (например, температура воздуха в районе МГТУ ГА, замеряемая датчиком с электронным табло) подчиняется некоторому закону распределения $F(\xi)$, то каждое наблюдаемое ее значение x_i в i -ый замер встречается с какой-то вероятностью, определяемой с помощью данного закона распределения.

В случае **дискретной** случайной величины (например, число очков, выпавшее на игральной кости) *закон распределения* задается таблицей соответствия возможных значений и вероятностей их появления (см. табл. 6).

Таблица 6.

Закон распределения				
ξ_i	ξ_1	ξ_2	...	ξ_M
$P(\xi_i)$	p_1	p_2	...	p_M

Для игральной кости						
x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Для непрерывных случайных величин такое соответствие записать нельзя, так как на любом отрезке числовой оси (на которой случайная величина может принимать значения) различных возможных значений бесконечно много и вероятность появления каждого из них равна нулю. Поэтому применяется запись закона распределения с помощью, так называемой интегральной функции распределения вероятностей $F(x) = P(\xi < x)$, задающей по определению **вероятность** того, что случайная величина ξ примет значение на числовой оси **левее** числа x . Понятно, что, чем правее расположена задаваемая граница x , тем больше вероятность попадания в соответственно больший интервал. Таким образом, выяснено главное свойство интегральной функции распределения $F(x)$ – **монотонное** возрастание ее значений от 0 до 1 на том интервале, на котором задана случайная величина. В общем случае следует говорить не о конечном интервале, а обо всей числовой оси, поэтому общий вид интегральной функции распределения $F(x)$ можно представить левым графиком рис. 43.

Справа на рис. 43 показана *дифференциальная функция распределения вероятностей* (*плотность распределения вероятностей*) $f(x)$, полученная дифференцированием интегральной $F(x)$. **Связь** этих функций представляется в виде:

$$\frac{dF(x)}{dx} = f(x); \quad F(x) = \int_{-\infty}^x f(x) \cdot dx.$$

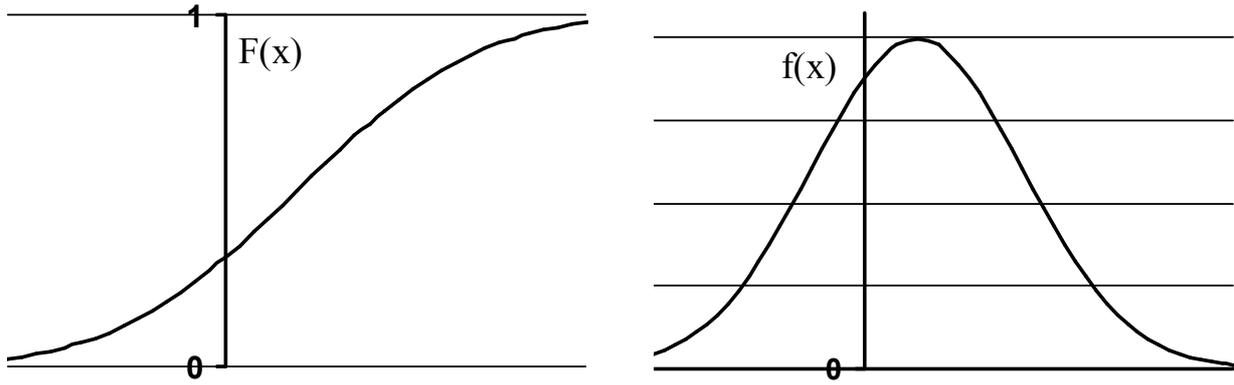


Рис. 43.

Исходя из определения интегральной функция распределения, для любого непрерывного закона распределения справедлива формула, определяющая вероятность попадания случайной величины в **интервал** от x_1 до x_2 :

$$P(x_1 < \xi < x_2) = F(x_2) - F(x_1).$$

В табл. 7 приводится система обозначений (для **дискретной** случайной величины ξ , принимающей M значений), принятая в данном учебном пособии. Обратим пока основное внимание на левую часть этой таблицы.

Таблица 7.

Простой случайный отбор

элементы генеральной совокупности	генеральные			элементы выборки	выборочные оценки		
	объем	средняя	дисперсия		объем	средняя	дисперсия
ξ_1, \dots, ξ_M	M	a	D, σ^2	x_1, \dots, x_N	N	\bar{x}	D_B, s^2

Расслоенный отбор

№ слоя	элементы генеральной совокупности	генеральные			элементы выборки	выборочные оценки		
		объем	средняя	дисперсия		объем	средняя	дисперсия
1	$\xi_{11}, \dots, \xi_{1M_1}$	M_1	a_1	D_1, σ_1^2	x_{11}, \dots, x_{1N_1}	N_1	\bar{x}_1	D_{B1}, s_1^2
2	$\xi_{21}, \dots, \xi_{2M_2}$	M_2	a_2	D_2, σ_2^2	x_{21}, \dots, x_{2N_2}	N_2	\bar{x}_2	D_{B2}, s_2^2
...
k	$\xi_{k1}, \dots, \xi_{kM_k}$	M_k	a_k	D_k, σ_k^2	x_{k1}, \dots, x_{kN_k}	N_k	\bar{x}_k	D_{Bk}, s_k^2

Здесь объемы связаны следующим образом: $M = \sum_{j=1}^k M_j$, $N = \sum_{j=1}^k N_j$, а индексация значений расслоенной случайной величины имеет вид: ξ_{ji} (или x_{ji}), где первый индекс означает номер слоя (группы) $j = 1, 2, \dots, k$, а второй – порядковый номер в слое $i = 1, 2, \dots, M_k$ (N_k).

Основными *числовыми характеристиками* законов распределения являются математическое ожидание, обозначаемое a , и дисперсия, обозначаемая D

или σ^2 . Величина σ имеет собственное наименование – среднее квадратическое отклонение.

Математическое ожидание $a \equiv E(\xi)$ характеризует центр закона распределения – ту точку на числовой оси, около которой следует ожидать появления случайной величины. Для дискретной случайной величины математическое ожидание определяется формулами:

$$a \equiv E(\xi) = \sum_{i=1}^M \xi_i p_i, \quad \text{или для } j\text{-го слоя: } a_j = \sum_{i=1}^{M_j} \xi_{ji} p_{ji},$$

здесь $E(\xi)$ обозначает операцию вычисления по приведенной формуле. Для непрерывной случайной величины математическое ожидание определяется формулой:

$$a \equiv E(\xi) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx.$$

Дисперсия σ^2 и среднее квадратическое отклонение σ характеризуют **рассеивание** (разброс) значений случайной величины около ее математического ожидания – возможные отклонения – "размазанность" кривой плотности распределения. Чем больше дисперсия, тем шире "колокол" этой функции. Для дискретной случайной величины дисперсия определяется формулами:

$$\sigma^2 \equiv D(\xi) \equiv E(\xi - a)^2 = \sum_{i=1}^M (\xi_i - a)^2 p_i, \quad \sigma_j^2 = \sum_{i=1}^{M_j} (\xi_{ji} - a_j)^2 p_{ji},$$

а для непрерывной случайной величины – формулой:

$$\sigma^2 \equiv D(\xi) = \int_{-\infty}^{\infty} (x - a)^2 \cdot f(x) \cdot dx.$$

Далее в данном учебном пособии будут рассматриваться такие дискретные случайные величины, которые принимают M различных значений с одинаковой вероятностью. В реальности таких случайных величин не бывает, однако для простоты изложения математического аппарата такое можно себе представить, тем более что именно так ведут себя результаты наблюдений и измерений. Для такой модели дискретных случайных величин все $p_i = \frac{1}{M}$ или $p_{ji} = \frac{1}{M_j}$ одинаковы, поэтому числовые характеристики задаются достаточно простыми формулами:

$$a \equiv E(\xi) = \frac{1}{M} \sum_{i=1}^M \xi_i, \quad a_j = \frac{1}{M_j} \sum_{i=1}^{M_j} \xi_{ji},$$

$$\sigma^2 \equiv D(\xi) \equiv E(\xi - a)^2 = \frac{1}{M} \sum_{i=1}^M (\xi_i - a)^2, \quad \sigma_j^2 = \frac{1}{M_j} \sum_{i=1}^{M_j} (\xi_{ji} - a_j)^2.$$

Существуют и другие числовые характеристики случайных величин. Среди таковых для центра распределения является медиана \tilde{x} – такое значение, для которого одинаково вероятно, окажется ли случайная величина меньше или больше \tilde{x} , а также мода x_M – наиболее вероятное значение случайной величины. Еще одной характеристикой рассеивания является размах R – разность между наибольшим и наименьшим из возможных значений случайной величины.

Для характеристики системы двух случайных величин ξ , η (встречающихся парами) используется ковариация (*корреляционный момент*) определяемая следующим образом:

$$\text{cov}(\xi, \eta) \equiv E(\xi - a, \eta - b) = \frac{1}{M} \sum_{i=1}^M (\xi_i - a)(\eta_i - b),$$

где b – математическое ожидание случайной величины η ; ξ_i, η_i – все возможные пары значений ξ, η . Нормирование ковариации по средним квадратическим отклонениям дает коэффициент корреляции:

$$\rho_{\xi\eta} = \frac{\text{cov}(\xi, \eta)}{\sigma_{\xi}\sigma_{\eta}}.$$

Указанные числовые характеристики **реальных** случайных величин нам не могут быть известны, так же как их объемы M, M_j , вероятности p_i, p_{ji} и законы распределения – это может быть известно лишь Создателю. В этих формулах $E(\bullet)$ и $D(\bullet)$ лишь обозначают операции **определения** математического ожидания и дисперсии, но по ним эти характеристики обычно **не вычисляются**. Эти формулы можно использовать только для **модельных** случайных величин, законы распределения которых записаны априори, как, например, для игральной кости.

Очевидно, что на практике не всегда есть возможность построить идеализированную схему для расчета вероятности по *классическому определению*. Например, для того, чтобы выявить поддельность игральной кости необходимо убедиться в неодинаковости вероятностей выпадения различного количества очков. В этом случае нет подходящей схемы равновозможных исходов для классического определения вероятностей. Поэтому приходится его заменять статистическим.

Статистическое определение вероятности основывается на априорном свойстве **состоятельности** любого массового повторения опытов. Это значит, что при бесконечном увеличении числа повторений опытов относительная частота появления интересующего нас события стремится к вероятности:

$$\frac{\mu}{v} \xrightarrow{v \rightarrow \infty} P(\mathcal{A}) = \frac{m}{n},$$

где μ – число появлений события \mathcal{A} в v опытах. Иначе говорят, что $\frac{\mu}{v}$ **сходится по вероятности** к величине $P(\mathcal{A})$: $P\left(\left|\frac{\mu}{v} - P(\mathcal{A})\right| < \varepsilon\right) \xrightarrow{v \rightarrow \infty} 1$. Величину

$\lim_{v \rightarrow \infty} \frac{\mu}{v}$ и принимают за статистическое определение вероятности. Разумеется, точно этот предел найти нельзя, но **оценить** его с той или иной уверенностью (**вероятностью**) на конечном множестве опытов можно. Этим и занимается математическая статистика.

Математическая статистика – наука для разработки методов регистрации, описания и анализа экспериментальных данных наблюдения массовых случай-

ных явлений. Центральное место в математической статистике занимают *теория оценок* и *теория проверки гипотез*. Основное правило математической статистики гласит: каждое выдвинутое предложение должно быть *оценено* и *проверено* на правдоподобие.

Всякое наблюдение дает какое-то ограниченное представление о явлении в целом, в том числе и об определенной случайной величине. Итогом такого наблюдения, т.е. отбора информации, становится выборка – некоторая совокупность результатов наблюдения случайной величины (генеральной совокупности), отобранная для исследования по определенному правилу. Таким образом, результаты наблюдения случайного явления дают лишь ограниченную информацию о случайной величине в целом. Поэтому переход от зафиксированного экспериментального факта к выводу и прогнозу – далеко не очевиден и нуждается в обосновании. Математическая статистика позволяет по результатам наблюдения **частного** (*выборки*) сделать некоторые обоснованные выводы о характеристиках **общего** (*генеральной совокупности*).

Вся продукция электролампового завода – генеральная совокупность, проверенная часть ламп – выборка. По сроку службы в **среднем**, вычисленному по выборке, можно судить о сроке службы ламп в **среднем** для всей генеральной совокупности. Но лишь "судить", "оценивать", ибо нельзя знать заранее срок службы какой-либо конкретной лампы.

Интервалы времени между прибытием самолетов в аэропорт 13 февраля с 12⁰⁰ до 16⁰⁰ дают какое-то представление о том, что будет 14 февраля с 12⁰⁰ до 16⁰⁰ – например, в каком режиме придется работать АТБ.

Общая последовательность применения методов математической статистики была предложена Р. Фишером (в скобках дается комментарий в современных терминах):

- 1▶ Планирование исследований (*планирование эксперимента*, определение способа отбора информации).
- 2▶ Конкретизация математико-статистического описания (выбор *дисперсионной* или *регрессионной модели*).
- 3▶ Оценка параметров модели (получение *точечных* и *интервальных оценок*) и составление их выборочных (эмпирических) распределений.
- 4▶ Изучение согласия между моделью и наблюдениями (*адекватность модели* оригиналу и проверка критериев согласия в обоснование модели).
- 5▶ Реальное решение задачи посредством оценок параметров и критериев значимости (*статистический анализ* результатов и разработка выводов).

Первый и третий этапы представляют собой, так называемую, процедуру первичной обработки информации, которая начинается с отбора информации, включает построение гистограмм и полигонов частот (если это необходимо для визуальной оценки вида эмпирического распределения), и завершается расчетом точечных и интервальных **оценок**. Эти результаты служат исходным материалом для четвертого и пятого этапов – статистического анализа, целью которого является установление статистических **закономерностей**.

5.2. Отбор информации

Исследователь, озабоченный составлением некоторой модели явления, стремится получить из эксперимента лишь ограниченный круг параметров, ее описывающих. Под эту задачу и собирается информация. Таким образом формулируются первичные требования к отбору информации.

Однако, как мы убедимся на примере в конце параграфа, статистический материал может содержать в себе намного больше информации, чем ставилось целью собрать. Можно не только получить значения наблюдаемых параметров с контролируемой погрешностью и с заданной вероятностью, что позволяет сделать первичная обработка информации. Умелое извлечение информации позволяет еще и быстрее достичь требуемого результата, и оценить его добротность, и даже оценить адекватность разрабатываемой модели.

Но не стоит думать, что математическая статистика отвечает на все вопросы. Во-первых, она способна обработать только ту информацию, которая собрана. Во-вторых, ее методы дают лишь более или менее вероятные (и эту вероятность можно оценить) результаты. Поэтому нет и не может быть математической обработки информации **вообще**, есть только аппарат для целевых исследований, невозможных без предварительных предположений о модели. Т.е. за человеком в любом случае остается творческий подход к выбору модели, к способам отбора информации и к формулировке выводов.

Для анализа информации не важно, получена она из натурального эксперимента, или из вычислительного эксперимента на детерминированной или стохастической модели. В любом случае основными проблемами сбора и обработки информации являются:

- выбор существенных факторов;
- выбор процедуры отбора информации;
- обеспечение достоверности выводов по результатам анализа.

Все эти проблемы разрешимы с помощью математического аппарата статистического анализа, излагаемого в последующих параграфах.

Следующий пример иллюстрирует необходимость строгого научного подхода к сбору и обработке информации.

Социологический опрос. Пусть результаты ответов 50 человек на некоторый вопрос представлены рядом знаков "+" – "да" и "-" – "нет":

+ + - + + + + - + - + - - - + + - + - + - - + - + - + + + - - - - + - + + + - + + + - + - + - + +

Подсчитаем количество тех и других знаков: "+" встречается 29 раз, "-" – 21 раз. Т.е. ответов "да" на 38 % больше, чем "нет". Но подождем делать соответствующие практические выводы. Подсчитаем баланс знаков, стоящих на четных местах: "+" – 18, "-" – 7; и на нечетных: "+" – 11, "-" – 14. Насколько можно доверять общим итогам опроса в таком случае?

В каждой практической задаче можно выдвинуть множество гипотез о происхождении таких разногласий, например, в рассматриваемом примере можно предположить и психологические особенности разных групп респондентов, и недоброкачественный (непредставительный) отбор респондентов, и не-

добросовестность сборщиков информации и т.д. Можно ли оценить степень их влияния на итоговый результат? И как отобрать "хорошую" информацию?

Математическая статистика может дать ответы на все эти вопросы! Конечно, она не сформулирует причины обнаруженного разногласия и не даст непосредственного рецепта для отбора информации. Но она может оценить **вероятность** правильности выдвигаемых **гипотез**, определить число факторов, которые оказали решающее влияние на результаты, и оценить вклад каждого из них. С помощью математической статистики можно оценить и добротность самого статистического материала.

Отбор информации – важнейшая операция, от правильного проведения которой существенно зависит результат статистического анализа, а, следовательно, и выводы исследований. Не стоит заблуждаться относительно якобы объективности данных наблюдения или эксперимента. Во-первых, результаты таких наблюдений всегда имеют свойства случайной выборки из генеральной совокупности. Так обстоит дело и при контроле технологических процессов, когда проверяются не все характеристики не всех изделий; и при изучении природных явлений, когда не все факторы учитываются и контролируются; и при экспертизе, в которой принимает участие ограниченный круг экспертов. Во-вторых, информация собирается для определенных целей. Для проверки различных моделей необходима различная информация, подчас одна в другую не преобразуемая. В-третьих, результаты наблюдений фиксируются всегда с какой-то погрешностью: из-за методики измерения, измерительной аппаратуры, округлений и т.п. Опыт человечества, давно заметившего эти обстоятельства, привел к простейшему приему отбора информации: чем ее больше, тем лучше. В теории измерений это нашло свое выражение в методике многократного повторения опытов в идентичных условиях.

Однако не всегда есть возможность повторять опыты: это может быть слишком дорогим удовольствием или в принципе невозможно. В связи с этим приходится подробнее рассматривать различные виды отбора информации в эксперименте – получения выборки.

Отбор информации, происходящий помимо воли исследователя, называется естественным в противоположность искусственному. Здесь не следует путать ситуацию с *пассивным и активным экспериментом* (§ 7.1) – естественный отбор предполагает получение информации в виде **констатации** определенных событий, процесс которой (констатации) не зависит от исследователя. Далее рассматриваются различные виды *искусственного отбора*.

Пристрастный отбор осуществляется по заранее намеченному признаку. Наука до XX века пользовалась именно пристрастным отбором информации: для выявления какой-либо зависимости в изучаемом природном явлении необходимо было избавиться от влияния "посторонних" факторов (например, от притяжения Земли, от проходящего трамвая и т.п.). Поэтому каждый отдельный опыт в эксперименте ставился в одних и тех же специальных условиях, имевших немаловажное значение и получивших наименование "чистоты эксперимента".

Случайный отбор производится с помощью случайных чисел по любой методике.

Механический отбор – отбор данных из всей совокупности по какому-либо правилу (например, каждый пятый).

Типический отбор – отбор из слоев (частей) всей имеющейся совокупности. Так делается отбор материала из отдельных партий продукции для технического контроля.

Аритмический отбор – частный случай *типического* и *механического*, когда отбор производится из равных групп по правилу, например: из первой группы берется первый элемент, из второй – второй, и т.д.

Пропорциональный отбор – частный случай типического отбора, когда из каждого слоя отбирается часть, пропорциональная объему слоя.

При репрезентативном отборе получается представительная выборка, достаточно **полно** характеризующая **всю** совокупность с точки зрения влияния важных и существенных факторов. Безусловно, к такому отбору следует стремиться, однако для оценки степени репрезентативности необходимы именно те характеристики, которые являются результатом конечного анализа отобранной информации. Поэтому в таких областях, как, например, политическая социология или экология, где цена принимаемого решения чрезвычайно высока, проводится специальный статистический эксперимент для оценки репрезентативности различных выборок и построения оптимальной из них по определенному критерию.

Существуют и другие виды отбора информации, в частности, расслоенный случайный отбор – комбинация *типического* и *случайного*, при которой из отдельных слоев (групп, частей) отбор осуществляется случайным образом.

Объем расслоенных выборок может быть произвольным или регулироваться: пропорционально объему слоев или оптимально – для обращения, например, в минимум дисперсии результатов σ^2 или стоимости эксперимента C и т.д.

В правой части табл. 7 (стр. 103) приведена система обозначений для **выборочных** данных из генеральной совокупности, принятая в данном учебном пособии.

5.3. Точечные оценки

В процессе любого эксперимента, в том числе и вычислительного, приходится иметь дело с теми или иными значениями наблюдаемых параметров. Определение этих значений с достаточной точностью невозможно без многократных повторений опытов и специальной процедуры их обработки. Простейшим примером этого является определение средней величины результатов однотипных измерений. В более сложных случаях приходится вычислять значения ненаблюдаемых параметров по значениям наблюдаемых. Определение значения некоторого параметра наблюдаемого объекта по экспериментальным данным носит название статистической точечной оценки.

В первую очередь для расчетов и анализа любой случайной величины ξ необходимо получить точечные оценки **параметров закона** ее распределения, основными из которых являются *математическое ожидание* a и *дисперсия* σ^2 (см. § 5.1). Поэтому изучение *статистических методов точечных оценок* сосредоточено на получении оценок именно этих величин.

Наиболее простым методом нахождения точечных оценок является метод моментов, предложенный К. Пирсоном. Он заключается в приравнении начальных ν_r или центральных μ_r моментов порядка r генеральной совокупности соответствующим моментам выборки. Т.е. для оценки математического ожидания достаточно использовать формулу:

$$a \equiv \nu_1(\xi) \equiv \frac{1}{M} \sum_{i=1}^M \xi_i \quad \rightarrow \quad \bar{x} \equiv \nu_1(x) \equiv \frac{1}{N} \sum_{i=1}^N x_i,$$

где \bar{x} называется выборочной средней, а для оценки дисперсии – формулу:

$$\sigma^2 \equiv \mu_2(\xi) \equiv \frac{1}{M} \sum_{i=1}^M (\xi_i - a)^2 \quad \rightarrow \quad D_B \equiv \mu_2(x) \equiv \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2,$$

где D_B называется выборочной оценкой дисперсии.

Нетрудно заметить, что суммы в этих формулах отличаются друг от друга верхним пределом, что характеризует различие *классического* и *статистического* определений вероятности, и скобками: слева **все возможные** значения, справа **встретившиеся** в наблюдениях (в выборке). В этом переходе и содержится математический смысл метода моментов.

Основные преимущества метода моментов заключаются в простоте вычислений и **независимости** от законов распределения – их не нужно знать. Этим же методом можно построить точечные оценки для расслоенных выборок:

$$\bar{x}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_{ji}, \quad D_{Bj} = \frac{1}{N_j} \sum_{i=1}^{N_j} (x_{ji} - \bar{x}_j)^2,$$

$$\bar{\bar{x}} = \frac{1}{k} \sum_{j=1}^k \bar{x}_j = \frac{1}{N} \sum_{j=1}^k \sum_{i=1}^{N_j} x_{ji}, \quad D_B = \frac{1}{N} \sum_{j=1}^k \sum_{i=1}^{N_j} (x_{ji} - \bar{\bar{x}})^2.$$

Обратимся к табл. 7 и проследим процесс получения точечной оценки a математического ожидания (генеральной средней) для наблюдаемой в k выборках (расслоенный отбор) случайной величины. Даже если отвлечься от элементарных знаний математической статистики, ясно, что в качестве такой оценки может выступать выборочная средняя \bar{x} . Но если в нашем распоряжении k выборок (например, полученных k исследователями), то не ясно, какую из выборочных средних \bar{x}_j надо принять за оценку математического ожидания. Более того, эти выборочные средние \bar{x}_j сами по себе случайные величины, так как и выборки и условия экспериментов могут случайно меняться. А может быть вообще следует взять за искомую оценку какую-то другую величину, например, медиану, или моду, или среднее от средних? Тогда возникает вопрос, как это надо считать: ведь объем информации в разных выборках N_j различен!

Способ определения *точечной оценки* λ^* истинного значения параметра λ определяется теми ее свойствами, которые необходимо обеспечить в конкретном случае. Математическая статистика рассматривает следующие свойства точечных оценок.

Несмещенность – свойство точечной оценки λ^* , при котором ее математическое ожидание равно истинному значению λ оцениваемого параметра: $E(\lambda^*) = \lambda$. Т.е. несмещенная точечная оценка определяет искомый параметр **без систематической ошибки**: отклонения, например, \bar{x}_j от генеральной средней a (математического ожидания) распределены симметрично. В специальном курсе математической статистики доказывается, что выборочная средняя \bar{x} и исправленная выборочная оценка дисперсии s^2 обладают свойством несмещенности:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2,$$

$$s_j^2 = \frac{1}{N_j - 1} \sum_{i=1}^{N_j} (x_{ji} - \bar{x}_j)^2, \quad s^2 = \frac{1}{N-1} \sum_{j=1}^k \sum_{i=1}^{N_j} (x_{ji} - \bar{x})^2.$$

Состоятельность – свойство точечной оценки λ^* , при котором с возрастанием объема выборки N она **стремится** по вероятности к истинному значению λ оцениваемого параметра: $\lim_{N \rightarrow \infty} P(|\lambda^* - \lambda| < \varepsilon) = 1$, где P – вероятность, ε – произвольное сколь угодно малое число. В специальном курсе математической статистики доказывается, что выборочная средняя \bar{x} является состоятельной оценкой математического ожидания.

Эффективной называется точечная оценка λ^* , имеющая при заданном объеме выборки N **наименьшую дисперсию**. Естественно, что из нескольких *состоятельных* оценок следует выбирать наиболее эффективную. Препятствием выявления этого свойства может стать необходимость знания закона распределения изучаемой случайной величины. Поэтому далеко не всегда удается обеспечить эффективность найденной точечной оценки. Можно показать, что выборочное среднее \bar{x} является эффективной оценкой математического ожидания.

Достаточная (исчерпывающая) точечная оценка **не может** быть существенно изменена из-за получения какой-либо **дополнительной** информации. В этом смысле такая оценка обеспечивает полноту использования всей информации, содержащейся в выборке, и свидетельствует о том, что выборка репрезентативная (представительная). *Эффективная* оценка обязательно является *достаточной*.

Свойства оценок математического ожидания a с помощью выборочных оценок среднего \bar{x} , медианы \tilde{x} и моды x_M ; дисперсии σ^2 с помощью исправленной D_B и исправленной s^2 оценок дисперсии; а также среднего квадратического отклонения σ с помощью $\sqrt{D_B}$, s и размаха R сведены в табл. 8.

| Оцениваемый параметр | Вычисляемая характеристика | Свойства оценок | | |
|----------------------|----------------------------|-----------------|-----------------|----------------|
| | | несмещенность | состоятельность | эффективность |
| a | \bar{x} | + | + | + |
| a | \tilde{x} | – | + | – |
| a | x_M | – | + | – |
| σ^2 | $D_{\%o}$ | – | + | + |
| σ^2 | s^2 | + | + | + |
| σ | $\sqrt{D_{\hat{a}}}$ | – | + | + |
| σ | s | – | + | + |
| σ | R | – | + | + при $N < 10$ |

Метод моментов, изложенный выше, дает **состоятельные** оценки. Однако Р. Фишер показал, что полученные этим методом оценки могут быть **смещенными** и **неэффективными**. Он же разработал и обосновал другой метод, свободный от этих недостатков.

Метод наибольшего правдоподобия основывается на отыскании такой оценки параметра распределения λ^* , которая обращает в максимум вероятность появления именно той выборки (x_1, x_2, \dots, x_N) , которая получена в эксперименте. Для его реализации составляют функцию правдоподобия:

$$L = P(x_1, x_2, \dots, x_N, \lambda^*),$$

представляющую собой именно эту вероятность, определенную при искомом значении параметра λ^* закона распределения. (В общем случае может рассматриваться задача нахождения нескольких параметров, тогда под λ^* понимается вектор.) Далее тем или иным способом решается задача оптимизации – нахождения такого значения λ^* , которое обеспечивает функции правдоподобия наибольшее значение.

По виду функции правдоподобия ясно, что составить и вычислить ее можно только тогда, когда **известен вид** закона распределения. Только в этом случае можно вычислить при любом предполагаемом значении λ^* вероятность появления того набора значений случайной величины x_1, x_2, \dots, x_N , который получен в эксперименте.

Требование знания закона распределения и является практически единственным, но существенным недостатком метода. Однако никто не запрещает сравнивать между собой наибольшие значения функций правдоподобия, определенных для нескольких альтернативных видов распределения.

Поскольку в практике чаще всего используются законы распределения, описываемые с помощью экспонент, оказалось удобным находить критические точки функции правдоподобия из необходимых условий экстремума не для самой L , а для ее логарифма (что одно и то же для отыскания единственного максимума):

$$\frac{\partial \ln L}{\partial \lambda^*} = 0,$$

называемых уравнением наибольшего правдоподобия. В таком виде удобно применять метод еще и потому, что для **независимо** полученных значений случайной величины (а именно так и стремятся поставить эксперимент) функция правдоподобия принимает вид произведения:

$$L = P(x_1, x_2, \dots, x_N, \lambda^*) = P(x_1, \lambda^*) \cdot P(x_2, \lambda^*) \cdot \dots \cdot P(x_N, \lambda^*) = \prod_{i=1}^N P(x_i, \lambda^*),$$

что обеспечивает простоту уравнения наибольшего правдоподобия:

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{i=1}^N \frac{1}{P(x_i, \lambda^*)} \cdot \frac{\partial P(x_i, \lambda^*)}{\partial \lambda^*} = 0.$$

Применение метода наибольшего правдоподобия для нахождения оценки *математического ожидания* приводит к тем же результатам, что и метод моментов, что и неудивительно, так как в этом случае и метод моментов обеспечивал все требуемые свойства оценок.

Можно показать, что известный *метод наименьших квадратов* (см. § 6.3) является частным случаем метода наибольшего правдоподобия для оценки математического ожидания **нормально** распределенной случайной величины.

Необходимость получения состоятельных, эффективных и достаточных оценок в случаях **неизвестных** законов распределения привела к разработке приемов получения робастных оценок и критериев – не зависящих от вида закона распределения или, по крайней мере, устойчивых к его нарушениям. Однако обольщаться этими приемами не следует, так как они в основном имеют эмпирическое происхождение, связаны с удачными решениями определенных задач, а в общем случае теоретически **не обоснованы**, т.е. не гарантируют получения результата.

Перечисленные выше требования к оценкам наталкивают на вопрос о приемлемом для тех или иных оценок объеме информации. Более подробно этот вопрос освещен в § 7.3, здесь же достаточно упомянуть простейшие практические ограничения снизу на объем статистического материала для уверенных результатов: $N \geq 30$, $N_j \geq 5$.

Это ограничение касается не столько обеспечения состоятельности и эффективности оценки, сколько достаточности степеней свободы выборки. Числом степеней свободы для системы n случайных величин называется число n этих величин **минус** число линейных **связей** между ними. Поэтому при определении выборочного среднего – первой числовой характеристики (связи) – в качестве числа степеней свободы используется объем выборки N или, соответственно, N_j (число полученных в выборке случайных значений). При вычислении второй характеристики – дисперсии – число степеней свободы необходимо уменьшить на 1 за счет **использования** связи в виде выборочного среднего. Вот почему несмещенная оценка дисперсии требует в знаменателе своей фор-

мулы число $N - 1$ или $N_j - 1$. Кроме того, число степеней свободы, оправдывая свое название, указывает количество связей между величинами, которые **можно определить** на данном статистическом материале, не опасаясь их линейной зависимости и, соответственно, вырожденности результатов. Поэтому ограничение снизу $N_j \geq 5$ в 5 единиц оставляет свободу использования таких независимых характеристик, как среднее, дисперсия, медиана и мода.

В табл. 9 в обозначениях табл. 7 приведены формулы для вычисления требуемых объемов выборок из слоев, отобранных различными, описанными в § 5.2, способами, и выборочного среднего всей совокупности с его дисперсией.

Таблица 9.

| | Объем | Мат. ожидание | Дисперсия |
|--------------------------|--|---|---|
| Генеральная совокупность | | $\sum_{j=1}^k \mu_j a_j$ | $\sum_{j=1}^k \mu_j \sigma_j^2 + \sum_{j=1}^k \mu_j \left(a_j - \sum_{i=1}^k \mu_i a_i \right)^2$ |
| Расслоенный отбор | Требуемый N_j | Выборочное среднее | Дисперсия выб. среднего |
| Произвольный | | $\sum_{j=1}^k \mu_j \bar{x}_j$ | $\sum_{j=1}^k \mu_j^2 \frac{\sigma_j^2}{N_j} (1 - \alpha_j)$ |
| Пропорциональный | $\mu_i N$ | $\sum_{j=1}^k \mu_j \bar{x}_j$ | $\frac{1 - \alpha}{N} \cdot \sum_{j=1}^k \mu_j \sigma_j^2$ |
| Оптимальный (minD) | $\frac{\mu_i \sigma_i}{\sum_{j=1}^k \mu_j \sigma_j} \cdot N$ | $\sum_{j=1}^k \mu_j \sigma_j \cdot \sum_{j=1}^k \frac{\mu_j \bar{x}_j}{\sigma_j}$ | $\frac{\left(\sum_{j=1}^k \mu_j \sigma_j \right)^2}{N} - \frac{\sum_{j=1}^k \mu_j \sigma_j^2}{M}$ |
| Оптимальный (minC) | $\frac{\mu_i \sigma_i}{\sum_{j=1}^k \frac{\mu_j \sigma_j}{\sqrt{c_j}}} \cdot \frac{N}{\sqrt{c_i}}$ | $\sum_{j=1}^k \frac{\mu_j \sigma_j}{\sqrt{c_j}} \cdot \sum_{j=1}^k \frac{\mu_j \sqrt{c_j} \bar{x}_j}{\sigma_j}$ | $\frac{\left(\sum_{j=1}^k \frac{\mu_j \sigma_j}{\sqrt{c_j}} \right)^2}{N} \cdot \sum_{j=1}^k \mu_j c_j (1 - \alpha_j)$ |

Здесь использованы следующие обозначения: $\alpha = \frac{N}{M}$, $\mu_j = \frac{M_j}{M}$, $\alpha_j = \frac{N_j}{M_j}$;

стоимость эксперимента обозначена $C = C_0 + \sum_{j=1}^k c_j N_j$, где C_0 – накладные расходы, c_j – стоимость одного наблюдения в j -м слое.

5.4. Законы распределения

Фундаментальное значение в теории вероятностей имеет *центральная предельная теорема*, доказывающая, что сумма **произвольно** распределенных **независимых** случайных величин при условии **одинакового** их влияния распределена по *нормальному закону*. Нормальному закону распределения подчиняется, например, **случайная ошибка** измерений, к которой человечество привыкло за тысячелетия своей практической деятельности. На практике это означает хорошо известный на бытовом уровне факт, что погрешность измерений при их многократном повторении стремится к характерному колоколообразному распределению – *распределению ошибки* (см. рис. 43). Эта колоколообразная кривая, называемая кривой ошибок или кривой Гаусса, и представляет собой *плотность распределения вероятностей нормального закона*.

В общем виде нормальный закон распределения описывается функцией его плотности распределения вероятностей:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}$$

с математическим ожиданием a и дисперсией σ^2 и интегральной функцией:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-a}{\sigma}\right)^2} dt.$$

Поскольку интегральная функция распределения $F(x)$ используется в приложениях, а через элементарные функции не выражается, ее табулируют. Такую функцию для **стандартизованного** нормального закона при $a = 0$, $\sigma = 1$ называют функцией Лапласа и обозначают $\Phi(x)$. Через $\Phi(x)$ можно выразить интегральную функцию распределения для любых значений математического ожидания a и среднего квадратического отклонения σ . Существует несколько разновидностей функции Лапласа:

$$\Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad \Phi_1(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt,$$

$$F(x) = \Phi^*\left(\frac{x-a}{\sigma}\right) = 0,5 + \frac{1}{2} \Phi_1\left(\frac{x-a}{\sigma}\right) = 0,5 + \Phi\left(\frac{x-a}{\sigma}\right)$$

– $\Phi^*(x)$ затабулирована в [9], $\Phi_1(x)$ – в [26], $\Phi(x)$ – в [18].

В § 5.1 рассматривалось множество значений выборочных средних \bar{x}_j , претендовавших на оценку математического ожидания. Каждое из этих значений теперь можно рассматривать как отдельную реализацию некоторой не рассматривавшейся ранее случайной величины – выборочной средней (определяемой суммой случайных значений), которая, согласно центральной предельной теореме, распределена по **нормальному** закону. А так как закон распределения известен, то оценку математического ожидания генеральной совокупности

можно получить с заданными свойствами.

Эти рассуждения наводят на мысль об обоснованной возможности получения состоятельных и эффективных оценок не только математического ожидания, но и некоторых **функций** от него, используя такие величины, как среднее выборочное, выборочная оценка дисперсии, выборочная ковариация и т.д. Особенно важным это становится при обработке структурированных выборок большого объема (см. табл. 7), необходимых для статистического анализа.

В табл. 10 приведены распределения некоторых важных выборочных функций – функций от выборочных значений. Во всех этих случаях предполагается, что выборка объема N (или в слое N_j) сделана из нормально распределенной генеральной совокупности с математическим ожиданием a (или a_j) и дисперсией σ^2 (или σ_j^2). В табл. 10 обозначены следующие законы распределения, таблицы которых приводятся в специальной литературе: u – стандартизованное нормальное распределение с нулевым математическим ожиданием и единичной дисперсией (u -распределение); t – распределение Стьюдента (t -распределение); r – r -распределение; χ^2 – χ^2 -распределение Пирсона; F – распределение Фишера (v^2 -распределение); z – z -распределение. Кроме того введены обозначения: для характеристик расслоенных выборок:

$$s_A^2 = \frac{1}{k-1} \cdot \sum_{j=1}^k N_j (\bar{x}_j - \bar{\bar{x}})^2, \quad s_0^2 = \frac{1}{N-k} \cdot \sum_{j=1}^k (N_j - 1) s_j^2$$

– s_A^2 – *межгрупповая дисперсия* (выборочная оценка) между слоями (рассеяние из-за влияния исследуемого фактора) и s_0^2 – *остаточная (внутренняя) дисперсия* (выборочная оценка) внутри слоев (рассеяние результатов из-за влияния неучтенных факторов); а для системы случайных величин:

$$\beta_{\eta\xi} = \rho_{\xi\eta} \frac{\sigma_\eta}{\sigma_\xi}, \quad b_{yx} = r_{xy} \frac{s_y}{s_x},$$

$$\text{где } r_{xy} = r_{yx} = \frac{l_{xy}}{s_x s_y}, \quad a \quad l_{xy} = l_{yx} = \frac{1}{N} \cdot \sum_{j=1}^N (x_j - \bar{x})(y_j - \bar{y})$$

– гипотетический $\beta_{\eta\xi}$ (генеральный) и выборочный b_{yx} *коэффициенты регрессии*, выборочные *коэффициент корреляции и ковариация*.

Таблица 10.

| № | Выборочная функция | Закон распределения | Число степеней свободы закона |
|---|---|---------------------|-------------------------------|
| 1 | $\frac{x_i - a}{\sigma}$ | u | – |
| 2 | $\frac{x_i - a}{s} = \frac{x_i - a}{\sqrt{D_B}} \sqrt{\frac{N-1}{N}}$ | t | $N-1$ |

| № | Выборочная функция | Закон распределения | Число степеней свободы закона |
|----|--|---------------------|-------------------------------|
| 3 | $\frac{x_i - \bar{x}}{\sqrt{D_B}} = \frac{x_i - \bar{x}}{s} \sqrt{\frac{N}{N-1}}$ | r | N-1 |
| 4 | $\frac{\bar{x} - a}{\sigma} \sqrt{N}$ | u | – |
| 5 | $\frac{\bar{x} - a}{s} \sqrt{N} = \frac{\bar{x} - a}{\sqrt{D_B}} \sqrt{N-1}$ | t | N-1 |
| 6 | $\frac{ND_B}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \bar{x})^2$ | χ^2 | N-1 |
| 7 | $\frac{N\hat{s}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - a)^2$ | χ^2 | N |
| 8 | $(N-k) \frac{s_0^2}{\sigma^2}$ | χ^2 | N-k |
| 9 | $(k-1) \frac{s_A^2}{\sigma^2}$ | χ^2 | k-1 |
| 10 | $\frac{(\bar{x}_i - \bar{x}_j)(a_i - a_j)}{\sigma} \sqrt{\frac{N_i N_j}{N_i + N_j}}$ | u | – |
| 11 | $\frac{(\bar{x}_i - \bar{x}_j)(a_i - a_j)}{s_0} \sqrt{\frac{N_i N_j}{N_i + N_j}}$ | t | $N_i + N_j - 2$ |
| 12 | $\frac{s_i^2}{s_j^2}$ | F | $N_i - 1, N_j - 1$ |
| 13 | $\frac{s_A^2}{s_0^2}$ | F | k-1, N-k |
| 14 | $\sqrt{N-2} \frac{r_{xy}}{\sqrt{1-r_{xy}^2}}$ | t | N-2 |
| 15 | $r_{xy} \sqrt{N-1}$ | r | N-2 |
| 16 | $\frac{s_x \sqrt{N-2}}{s_y \sqrt{1-r_{xy}^2}} (b_{yx} - \beta_{\eta\xi})$ | t | N-2 |

В специальной литературе можно найти несколько более широкий список выборочных функций. Кроме того, в прикладных исследованиях можно поль-

зоваться не только точными законами распределения выборочных функций, но и приближенными. Перечень известных приближенных законов распределения выборочных функций значительно шире.

Таким образом, зная закон распределения выборочной функции, можно построить оценки наибольшего правдоподобия для параметров ее распределения. При этом математическое ожидание и дисперсия упомянутых в табл. 10 законов распределения принимают значения, приведенные в табл. 11, где под f понимается число степеней свободы соответствующего закона.

Таблица 11.

| Распределение | Мат. ожидание | Дисперсия |
|---------------|--------------------------------|--|
| u | 0 | 1 |
| t | 0 | $\frac{f}{f-2}$ ($f > 2$) |
| r | 0 | 1 |
| χ^2 | f | 2f |
| F | $\frac{f'}{f'-2}$ ($f' > 2$) | $\frac{2f'^2(f+f'-2)}{f(f'-2)^2(f'-4)}$ ($f' > 4$) |

Кроме того, законы распределения выборочных функций играют большую роль и в других задачах математической статистики, которые будут рассмотрены в следующих параграфах.

5.5. Интервальные оценки

Точечные оценки параметров распределения не всегда дают достаточно информации для анализа и выводов. Прежде всего, это связано с **приближенностью** полученных оценок из выборок небольшого объема. Кроме того, точечные оценки не дают информации об их точности. Использование для этой цели оценок дисперсии не всегда помогает, так как и они определяются неточно и имеют некоторую неединичную вероятность. Поэтому с некоторого момента стали использоваться интервальные оценки параметров закона распределения – доверительные интервалы. Идея Ю. Неймана об интервальных оценках заключается в получении некоторого интервала, в котором должен находиться оцениваемый параметр. Но, поскольку оценки такого рода делаются на основе случайной выборки, а не генеральной совокупности, постольку они должны даваться с определенной вероятностью. Так было сформулировано понятие доверительного интервала: интервал $(\lambda_1^*, \lambda_r^*)$ (λ_1 – левая граница, λ_r – правая граница), в котором с заданной *доверительной вероятностью* γ следует ожидать истинное, но не известное значение оцениваемого параметра λ , т.е.:

$$P(\lambda_1^* < \lambda < \lambda_r^*) = \gamma.$$

Выписанное соотношение является ключевым для всей процедуры отыскания доверительного интервала. Действительно: если известен закон распределения оцениваемого параметра λ (а его интегральная функция распределения по определению монотонно возрастает – см. вероятность попадания в интервал в § 5.1), то всегда можно подобрать множество пар значений $(\lambda_1^*, \lambda_r^*)$, удовлетворяющих определению доверительного интервала. Остается только договориться о конкретном выборе такой пары.

Естественным для такого выбора является опора на точечную оценку λ^* искомого параметра, найденную предварительно, и определение границ по этой величине:

$$\lambda_1^* = \lambda^* - \delta_l, \quad \lambda_r^* = \lambda^* + \delta_r,$$

где *погрешности* δ_l, δ_r (допуски) характеризуют **точность** оценки влево и вправо от λ^* . В простейшем случае принимают $\delta = \delta_l = \delta_r$, т.е. строят **симметричный** доверительный интервал относительно точечной оценки параметра.

Таким образом, оговорив соотношение между δ_l и δ_r , можно определить доверительный интервал однозначно, если только известен *закон распределения для выборочной функции* от этого параметра.

Найти симметричный доверительный интервал для математического ожидания a нормально распределенной случайной величины ξ в случае известного среднего квадратического отклонения σ .

В табл. 10 есть две выборочные функции, которые содержат искомый параметр a и известный параметр σ : в 1-й и в 4-й строках. Согласно принятой системе обозначений, с помощью первой строки можно определить a , исходя из единственного замера искомого параметра x_i , а с четвертой – исходя из выборочного среднего \bar{X} по выборке объемом N . Для этого проведем простейшие алгебраические преобразования на примере x_i :

$$P(x_i - \delta < a < x_i + \delta) = P(-\delta < a - x_i < \delta) = P\left(-\frac{\delta}{\sigma} < \frac{x_i - a}{\sigma} < \frac{\delta}{\sigma}\right).$$

При известном законе распределения величины $\frac{x_i - a}{\sigma}$ вычисление этой вероятности не представляет труда – достаточно воспользоваться формулой вероятности попадания в заданный интервал. В нашем случае эта величина согласно 1-й строке табл. 10 распределена по стандартизованному нормальному закону, таблицу функции Лапласа для которого возьмем из [18]. Для такой симметричной относительно нуля функции Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

выражение для определения доверительного интервала приобретает вид:

$$P\left(-\frac{\delta}{\sigma} < \frac{x_i - a}{\sigma} < \frac{\delta}{\sigma}\right) = \gamma = F\left(\frac{\delta}{\sigma}\right) - F\left(-\frac{\delta}{\sigma}\right) = 0,5 + \Phi\left(\frac{\delta}{\sigma}\right) - 0,5 - \Phi\left(-\frac{\delta}{\sigma}\right) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

При заданном γ по таблице функции Лапласа можно определить ее аргумент $u_{0,5\gamma} = \delta/\sigma$. Таким образом численное значение погрешности δ для точечной оценки математического ожидания a по данным единственного замера x_i определится: $\delta = u_{0,5\gamma} \sigma$, где $u_{0,5\gamma}$ – аргумент функции Лапласа, соответствующий значению функции, равному $\Phi(u_{0,5\gamma}) = 0,5 \cdot \gamma$. В итоге доверительный интервал приобретает вид:

$$x_i - u_{0,5\gamma} \sigma < a < x_i + u_{0,5\gamma} \sigma.$$

Аналогичным образом строится доверительный интервал для точечной оценки математического ожидания a по выборочному среднему \bar{x} из 4-й строки табл. 10, в результате чего, как нетрудно проверить, получается выражение:

$$\bar{x} - u_{0,5\gamma} \frac{\sigma}{\sqrt{N}} < a < \bar{x} + u_{0,5\gamma} \frac{\sigma}{\sqrt{N}},$$

свидетельствующее об уменьшении погрешности δ (увеличении точности) в \sqrt{N} раз по сравнению с единственным замером. Этот факт давно известен человечеству: "семь раз отмерь – один отрежь".

Этот пример разобран так подробно, чтобы показать возможность **получения** формулы доверительного интервала без использования специальных справочников для любого оцениваемого параметра с помощью таблицы законов распределения выборочных функций типа табл. 10. Так вычисляются доверительные интервалы для известных из курса математической статистики случаев: математического ожидания при неизвестном среднем квадратическом отклонении (из 5-ой строки), а также для дисперсии (из 6-ой).

5.6. Проверка статистических гипотез

В многообразной практической деятельности человека процедура выдвижения и проверки гипотез имеет самые различные формы от простейших бытовых обсуждений (насколько разбавлена сметана недобросовестным продавцом) до расчетов стоимости эффекта и потерь (какова средняя стоимость ремонта аварийного автомобиля – для страховой компании). Наиболее обоснованной из этих форм является математическая теория проверки статистических гипотез.

В математической статистике существует жесткое правило: любое предположение о свойствах распределения некоторой величины, основанное на **выборочных** данных, должно быть проверено. Статистическая проверка не может доказать истинность (чего не может сделать вообще никто), но может указать с некоторой долей уверенности на **наличие или отсутствие признаков опровержения** данного суждения. Дальше уже дело человека – принять или отвергнуть предлагаемую гипотезу на основании такого статистического вывода.

Некоторые вопросы статистической проверки затрагивались при изучении свойств точечных оценок. Например, неприемлемость в некоторых приложениях смещенной оценки дисперсии проистекало именно из-за того, что возникали признаки опровержения полученных результатов.

Собственно проверка статистических гипотез представляет собой аппарат получения оценки **соответствия** выдвинутой гипотезы полученному статистическому материалу, т.е. **выборке**.

В качестве выдвигаемых **гипотез** обычно выступают предположения о свойствах закона распределения $F(x, \lambda)$ генеральной совокупности. Таковыми могут быть предположения о значениях параметров λ закона распределения – для их проверки применяются *параметрические критерии*, а также предположения о непараметризуемых свойствах распределения – для них служат *непараметрические критерии*. Первые из таких гипотез принято обозначать $H_0: \lambda = \lambda_0$, вторые более общим выражением $H_0: F(x) = F_0(x, \lambda_0)$. Первые требуют знания общего вида закона распределения, зато вторые оказываются менее эффективными. Под λ_0 здесь понимается вполне определенное число, на совпадение с которым желательно проверить параметр λ .

В качестве **критерия** соответствия понимается достижение определенного значения *функции правдоподобия* (см. § 5.3) полученной выборки. Т.е. если выборка попадает в область малого правдоподобия, то присутствуют признаки опровержения гипотезы – **есть основания отвергнуть** эту гипотезу. В противоположном случае **нет оснований отвергнуть** гипотезу.

Однако, поскольку наши суждения о реальности весьма ограничены, постольку нельзя забывать о возможной ошибке в наших выводах, поэтому здесь возможны не два, а четыре исхода:

- 1) гипотеза **верна** и **не отвергается** согласно критерию (правильный вывод);
- 2) гипотеза **неверна** и **отвергается** согласно критерию (правильный вывод);
- 3) гипотеза **верна**, но **отвергается** согласно критерию (ошибка I рода);
- 4) гипотеза **неверна**, но **не отвергается** согласно критерию (ошибка II рода).

Вероятность ошибки I рода, т.е. вероятность ошибки **при условии** верности гипотезы, принято обозначать α и называть уровнем значимости критерия. Для оценки гипотезы необходимо **назначать уровень значимости** – максимальное значение вероятности, которое принимается за **практическую невозможность** получения конкретной выборки с гипотетическими свойствами. Тогда вероятность правильного неотвергания проверяемой гипотезы (1-го исхода) равна $1 - \alpha$.

Вероятность ошибки II рода, т.е. вероятность ошибки **при условии** неверности гипотезы, может быть оценена только после конкретизации *альтернативной (конкурирующей) гипотезы*, поскольку в этом случае решающее значение имеет "отдаленность" исходной гипотезы от ее альтернативы. Эту вероятность принято обозначать β . При этом вероятность правильного отвергания проверяемой гипотезы (2-й исход) равна $1 - \beta$, эта величина называется мощностью критерия.

Рассмотрим более подробно аппарат проверки статистических гипотез на примере параметрических критериев для проверки гипотез о значении математического ожидания. Итак, в этом случае речь идет о проверке гипотезы $H_0: a = a_0$. Для дальнейших рассуждений нам понадобится вид функции плотности распределения $f(\bar{x}, a) = f[\varphi(x_1, x_2, \dots, x_N), a]$ вероятности появления **наблюдаемого** значения выборочной функции в **предположении** равенства математического ожидания a значению a_0 (в соответствии с выдвинутой гипотезой). Пусть такое распределение, полученное с помощью табл. 10, изображено на рис. 44. В нашем примере такой выборочной функцией выступает среднее выборочное всех наблюдаемых результатов: $\bar{x} = \varphi(x_1, x_2, \dots, x_N) = \frac{1}{N} \sum_{i=1}^N x_i$. (Для оценки других параметров следует использовать соответствующие выборочные функции, содержащие вместо \bar{x} оцениваемый параметр, принципиальные рассуждения остаются теми же.) Тогда вероятность ошибки I рода α распределится на две в общем случае неравные части α_1 и $\alpha - \alpha_1$, которым соответствуют две части *критической области*, заштрихованной на рис. 44. Так как погрешность определения математического ожидания обычно рассматривается одинаковой в отрицательную и положительную стороны, то обе части критической области будем считать начинающимися на равном удалении от оценки a_0 . Средняя часть значений \bar{x} , вне критической области, соответствует **допустимым** оценкам математического ожидания, которые могут считаться **неотличимыми** от принятой a_0 . Такая оценка может, например, быть получена вычислением выборочного среднего или иным способом.

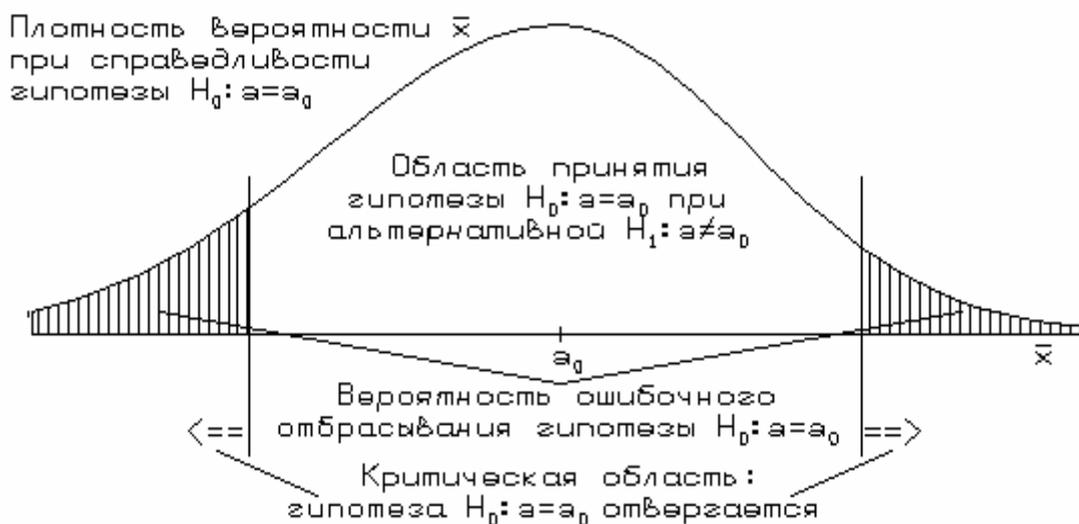


Рис. 44.

Если полученное в эксперименте выборочное среднее \bar{x} (вычисленное в качестве оценки математического ожидания a_0) попадет в критическую область, то это будет означать, что **различие** между \bar{x} и a_0 **значимо** (существенно, неслучайно, проистекает от каких-то неучтенных факторов, им нельзя пренебречь), и гипотеза $H_0: a = a_0$ **не может быть принята**.

Если выборочное среднее \bar{x} не попадет в критическую область, то это будет означать, что **различие** между \bar{x} и a_0 незначимо (несущественно, случайно, им можно пренебречь) и **нет оснований отвергнуть** гипотезу $H_0: a = a_0$, она может быть принята.

Проведенное разбиение критической области на две части на самом деле подразумевает введение дополнительного момента в наши построения: предположения о том, что отклонение от a_0 **в любую сторону на любую величину** одинаково неприемлемо. Т.е. вводится вторая – *альтернативная (конкурирующая) гипотеза* вида $H_0: a \neq a_0$, которая не оговаривает направление отклонения. Именно применение такой альтернативы позволяет конкретизировать вид и расположение критической области, как симметричной – "двухсторонней". В этом случае *критерий* принято называть также *двухсторонним*.

Рис. 45 иллюстрирует ситуацию с альтернативной гипотезой $H_0: a < a_0$, а рис. 46 – с $H_0: a > a_0$. В этих случаях дополнительных пояснений не требуется.

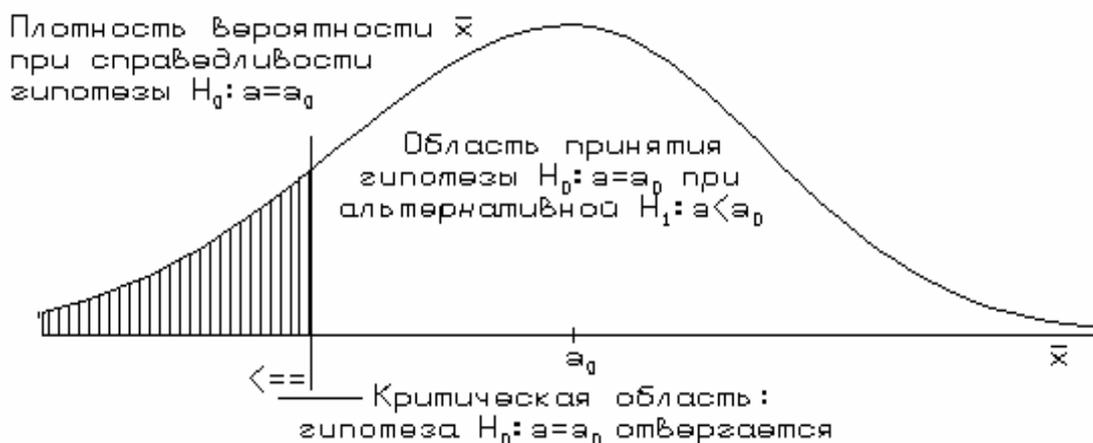


Рис. 45.

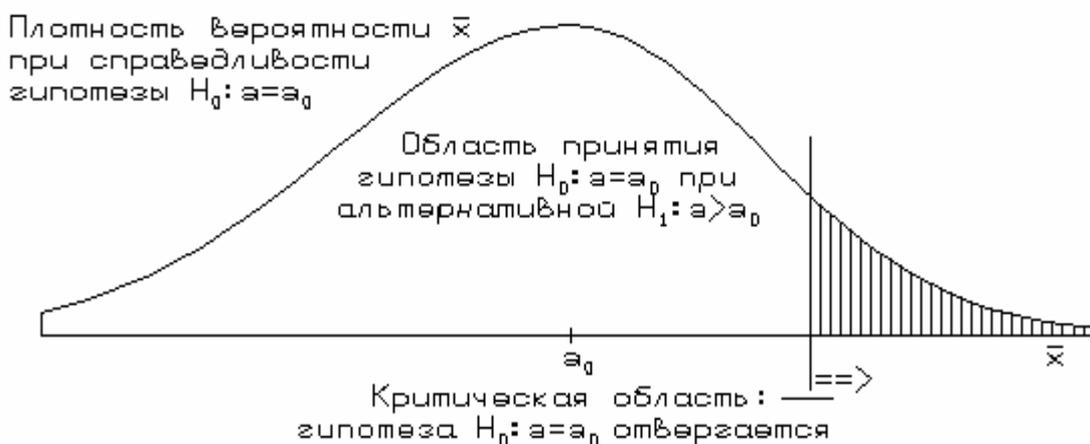


Рис. 46.

Но возможен еще один вид альтернативной гипотезы: $H_1: a = a_1$, когда фактически ставится задача оценки одного варианта из двух, различающихся на некоторую величину $a_1 - a_0$. Этот случай мы рассмотрим подробнее с помощью рис. 47, на котором для определенности $a_1 > a_0$.

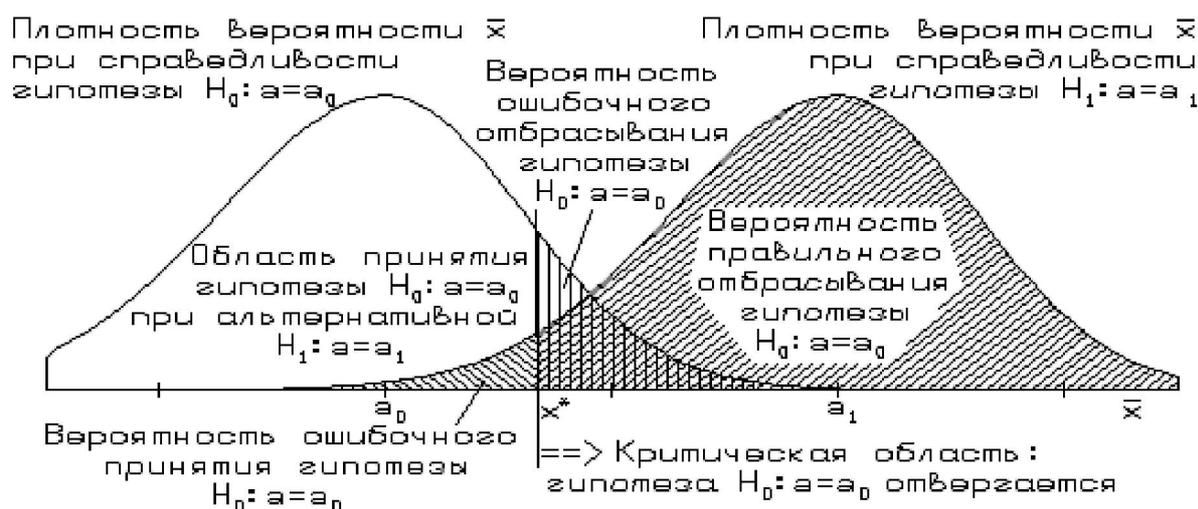


Рис. 47.

Как и прежде, граница критической области проводится по распределению, соответствующему исходной гипотезе, исходя из заданного уровня значимости α . При попадании \bar{x} в нее (правее границы x^*) исходную гипотезу $H_0: a = a_0$ следует отвергнуть, а левее – нет таких оснований. На рис. 47 вероятности α ошибки I рода соответствует область **под левой** кривой **правее** границы x^* критической области. Однако аналогичные рассуждения можно провести с **той же** границей x^* для альтернативной гипотезы $H_1: a = a_1$ при заданном значении β . Вероятность β ошибки II рода соответствует области **под правой** кривой **левее** границы критической области x^* . Очевидно, что в этом случае выбирать произвольно три величины: α , β и $a_1 - a_0$ нельзя. Формулировка альтернативной гипотезы необходима для конкретизации практической задачи и для разумного выбора α и β , исходя из оценки важности каждого из четырех возможных исходов. Эта графическая интерпретация помогает при анализе возможных исходов и **выборе** параметров α и β .

Практическая реализация проверки статистических гипотез осуществляется с помощью вычисления критерия значимости, представляющего собой некоторую из выборочных функций, закон распределения которой известен. Таким образом, алгоритм проверки статистических гипотез с помощью параметрических критериев может выглядеть следующим образом:

- 1 ► выдвижение оцениваемой гипотезы H_0 ;
- 2 ► выдвижение альтернативной гипотезы H_1 ;
- 3 ► установление подходящего уровня значимости α ;
- 4 ► выбор подходящей выборочной функции по следующим признакам:
 - подчиненность известному закону распределения (хотя бы с контролируемым **приближением**),
 - простота вычислений,
 - обеспечение наилучшего критерия (наиболее мощного: $1 - \beta \rightarrow 1$);
- 5 ► определение (вычисление или построение) распределения используемой выборочной функции в предположении гипотезы H_0 ;

6► определение критической области для проверки гипотезы H_0 с учетом альтернативной H_1 ;

7► получение выборки и вычисление значения выборочной функции (вычисление "статистики");

8► принятие решения: если вычисленное в предыдущем пункте значение ("статистика") попало в критическую область, то гипотезу следует отвергнуть, в противном случае нет оснований отвергнуть гипотезу.

Однако есть специальные случаи, когда указанный алгоритм не имеет ярко выраженного заверщенного вида.

В последнем из рассмотренных случаев альтернативных критериев несогласованность задаваемых параметров α , β и $a_1 - a_0$ может приводить к такой ситуации, когда невозможно сделать определенный вывод. Тогда, если задаваемые параметры α , β и $a_1 - a_0$ выбраны на основании физического смысла и требований, то указанное неопределенное состояние может свидетельствовать только об одном: статистическая информация не соответствует поставленной задаче исследований. В этом случае ее необходимо **обновлять или расширять**.

Подобные вопросы возникают и в случае невозможности обосновать априорное значение σ . Простейший подход к этой проблеме состоит в построении специального, **отдельного** эксперимента для нахождения его оценки σ , степени ее достоверности и значимости. Только после этого можно применять математический аппарат для решения практической задачи. Но здравый смысл подсказывает, что в принципе должна быть возможность объединить такую сугубо предварительную процедуру с собственно экспериментом. Чисто механическое объединение здесь, конечно, невозможно: достаточно вспомнить (см. § 5.3), что число степеней свободы при вычислении статистической оценки σ существенно зависит от количества проведенных опытов (наблюдений) и способов определения **других** точечных характеристик распределения.

Такая ситуация характерна для *последовательного анализа* и применения *секвенциальных* (последовательных) *критериев*. Ниже рассматривается один пример такого критерия – критерий Вальда. Однако применение последовательного анализа следует рассматривать значительно шире: на каждой стадии эксперимента необходимо оценить его результаты с точки зрения ответа на поставленный вопрос и, если уверенного ответа ни в положительном, ни в отрицательном смысле не удастся получить, **продолжать** эксперимент.

Для выбора между двумя гипотезами $H_0: a = a_0$ и $H_1: a = a_1$, где $a_1 > a_0$, задаются вероятности ошибочного их отвергания: α и β , соответственно. Тогда с помощью *секвенциального* (последовательного) *критерия* можно сделать один из **трех** выводов: принять проверяемую гипотезу, принять альтернативную гипотезу, продолжить эксперимент для увеличения объема выборки. При соблюдении условий: вид функции распределения известен, гипотезы фиксированы, α и β выбраны – А. Вальд предложил секвенциальный критерий отношения вероятностей:

$$\frac{\beta}{1-\alpha} < \frac{p_{1m}}{p_{0m}} < \frac{1-\beta}{\alpha},$$

где p_{0m} и p_{1m} – плотности вероятностей состоявшихся m наблюдений, вычисленные по функциям распределения с параметрами a_0 и a_1 соответственно. Если это неравенство выполняется, то следует продолжать эксперимент. При выходе отношения вероятностей за левую границу следует принять основную гипотезу $H_0: a = a_0$, при выходе за правую границу следует принять альтернативную гипотезу $H_1: a = a_1$. Алгоритмы секвенциальных критериев **всегда** заканчиваются принятием гипотез, причем требуют, как правило, вдвое меньшего объема эксперимента, чем классические критерии для оценки тех же гипотез (см. табл. 17).

Необходимо отметить, что на практике последним двухсторонним неравенством пользоваться не очень удобно: для малых α и β левое отношение много меньше 1, а правое – много больше. Поэтому отношение вероятностей выражают непосредственно через проверяемые законы распределения и алгебраическими преобразованиями сводят к неравенству для \bar{x}_m – выборочному среднему параметра x по результатам состоявшихся m наблюдений. Границы критерия для этой величины зависят только от m , причем с ростом m сближаются, как показано на рис. 48, обеспечивая сходимость метода. Для этой иллюстрации принято: $a_1 - a_0 = 1$, $\sigma = 3,5$, $\alpha = \beta = 0,05$.

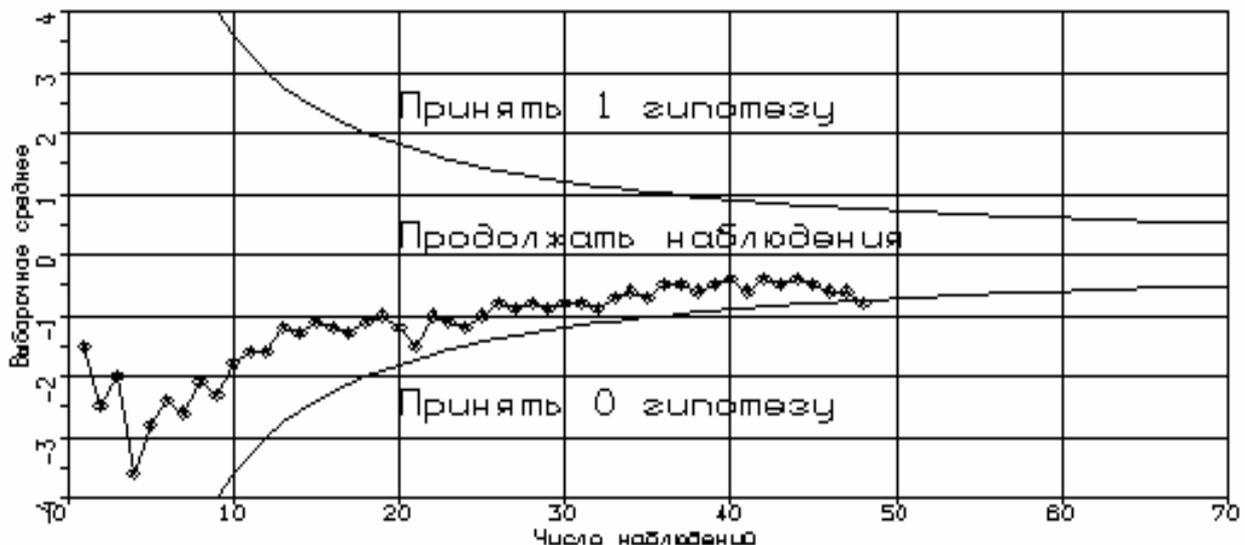


Рис. 48.

Все рассмотренные выше методы проверки гипотез относились к случаям оценки **параметров** выборочного **закона** распределения, вид которого предполагался известным хотя бы приближенно. В случае неизвестного закона распределения параметрические критерии не могут дать ответа на вопрос о справедливости статистических гипотез, поэтому применяются *непараметрические критерии*. Гипотезы, которые можно проверить с помощью непараметрических критериев, не касаются каких-либо числовых значений, а носят обобщенный характер, позволяющий применять их в областях, далеких от математики.

Основной задачей, где применяются непараметрические критерии, является задача сравнения двух совокупностей результатов: эмпирических с теоретическими или двух связанных или независимых эмпирических между собой. Если параметрические критерии для такого сравнения оперировали с числовыми характеристиками и известными законами распределения, то непараметрические критерии опираются на ранги – специфические числовые характеристики **упорядоченных** результатов эксперимента. Под упорядочением можно понимать и математические: больше – меньше, и нематематические классификационные признаки.

Простейший ранговый критерий – критерий знаков – использует только два ранга: да – нет или больше – меньше. Характерная двухранговая ситуация встречается в результатах опросов или тестирования. Здесь мы сформулируем критерий знаков для сопоставления двух **непрерывных** случайных величин ξ и η . Пусть имеются парные выборки одного объема N , в которых значения случайных величин ξ и η встречаются парами: $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$. В силу их непрерывности вероятность одинаковых значений в паре равна нулю, поэтому случаи практического совпадения из-за неточностей регистрации или округления отбрасываются. Если проверять гипотезу об **одинаковом распределении** величин ξ и η , как независимых, то должны совпадать вероятности:

$$P(x_i > y_i) = P(x_i < y_i) = 0,5.$$

Вероятность того, что среди этих N пар более m имеют положительные разности $x_i - y_i > 0$, легко подсчитывается в теории вероятностей через число сочетаний из N по j :

$$p_N(m) = \frac{1}{2^N} \sum_{j=m+1}^N C_N^j = \frac{1}{2^N} \sum_{j=m+1}^N \binom{N}{j}.$$

(В специальной литературе можно найти таблицы этой "функции накопления" биномиального распределения.) По заданному (выбранному) уровню значимости α определяется $m(\alpha)$ – **наименьшее** значение m , при котором $p_N(m) \leq \alpha$.

Если альтернативной является гипотеза о том, что случайная величина $\xi > \eta$, то критерий знаков выглядит следующим образом: для опровержения исходной гипотезы необходимо, чтобы **число** положительных разностей $x_i - y_i > 0$ было больше $m(\alpha)$ – т.е. в этом случае превосходство значений x_i над y_i неслучайно – **зна́чимо**. (При альтернативной гипотезе $\xi < \eta$ следует просто поменять случайные величины местами.)

Если альтернативной является гипотеза о том, что случайная величина ξ существенно отличается от случайной величины η в любую сторону ($\xi \neq \eta$), то критерий знаков требует отвергнуть исходную гипотезу, если **число** положительных **или** **число** отрицательных разностей окажется больше $m(\alpha)$ – при заданном уровне значимости 2α .

Этот же критерий знаков можно использовать для проверки гипотезы о значении медианы \tilde{x} эмпирического распределения: достаточно рассматривать разности полученных выборочных значений случайной величины и гипотети-

ческого значения медианы. Он применяется и для проверки симметрии закона распределения.

Другим из наиболее известных непараметрических критериев является критерий согласия К. Пирсона χ^2 для сравнения законов (не важно какого вида!) распределения двух случайных величин. Покажем его применение при проверке подчиненности полученной экспериментальной выборки определенному теоретическому (гипотетическому) закону распределения.

Весь диапазон N данных эксперимента разделим на r интервалов таким образом, чтобы обеспечить приемлемое число степеней свободы, т.е. число попаданий в i -й интервал наблюдаемых значений $N_i \geq 5$ (см. § 5.3). По этому же принципу число таких интервалов должно быть не менее 6, т.е. $N \geq 30$. Тогда наблюдаемое значение критерия Пирсона χ^2 определяется формулой:

$$\chi_{\text{наблюдаемое}}^2 = \sum_{i=1}^r \frac{(N_i - Np_i)^2}{Np_i}$$

где p_i – вероятность попадания в i -й интервал случайной величины, вычисленная по проверяемому теоретическому закону распределения. По специальной таблице распределения χ^2 при уровне значимости α (вероятности совершить ошибку первого рода: отвергнуть верную гипотезу) с $n = r - 2$ степенями свободы определяется величина $\chi_{\text{крит.}}^2(\alpha, n)$.

Если $\chi_{\text{наблюдаемое}}^2 < \chi_{\text{крит.}}^2(\alpha, n)$, то различие статистического и гипотетического законов распределения **незначимо**. Т.е. при заданном уровне значимости α нет оснований отвергнуть гипотезу о совпадении законов распределения, т.е. можно принять гипотезу о подчиненности экспериментальных данных гипотетическому закону распределения. В случае противоположного неравенства: $\chi_{\text{наблюдаемое}}^2 > \chi_{\text{крит.}}^2(\alpha, n)$ расхождение **значимо** (не может считаться случайным) и гипотезу следует отвергнуть, т.е. результаты эксперимента не описываются гипотетическим законом распределения.

5.7. Статистическая проверка адекватности математических моделей

В § 2.2 разработан общий подход к оценке адекватности математических моделей механических систем и процессов. Для адекватности математической модели реальному поведению оригинала рассогласование соответствующих параметров должно удовлетворять двум критериям: **точности** и **непротиворечивости**. Таким образом, необходим алгоритм проверки этих двух критериев для величины **рассогласования** результатов *контрольного вычислительного эксперимента* с результатами натурального эксперимента в тех же условиях: $\Delta u = u_{\text{модели}} - U_{\text{оригинала}}$. Эта величина на практике принимает дискретные значения, так как данные об оригинале регистрируются аппаратурой в конечном числе точек замера. Задачей контрольного вычислительного эксперимента в

этом случае является получение данных о тех же параметрах в тех же узловых точках. В результате для статистического анализа предлагается множество значений рассогласования $\{\Delta u_i; i = 1, 2, \dots, N\}$ – выборка из генеральной совокупности истинной величины рассогласования.

Как известно даже на бытовом уровне, для повышения **точности** измерений проводят не одно измерение, а несколько. Это делается не из-за того, что какое-то из них может оказаться ошибочным, а из-за замечательного свойства *дисперсии средней арифметической величины* измерений: уменьшаться с ростом числа повторений опытов:

$$D_N = \frac{D}{N}, \quad \text{или} \quad \sigma_N = \frac{\sigma}{\sqrt{N}},$$

где D и σ – дисперсия и среднее квадратическое отклонение в одном опыте (измерении), D_N и σ_N – дисперсия и среднее квадратическое отклонение результата осреднения замеров по N опытам. В справедливости этих формул нетрудно убедиться на основании математических определений указанных величин (см. § 5.1). Поэтому с помощью бóльшего числа опытов достигают меньшего рассеивания (среднего квадратического отклонения) данных, т.е. большей точности.

Однако одной величины *среднего квадратического отклонения* для оценки точности результатов недостаточно. Такая оценка страдает неполнотой, так как не учитывает, насколько часто встречаются большие и малые, положительные и отрицательные рассогласования.

Итак, точность следует определять единой оценкой всего множества наблюдаемых значений случайной величины рассогласования результатов вычислительного эксперимента и "истинного" значения наблюдаемой величины. По своему смыслу в качестве такой "истинной" единой оценки должно выступать *математическое ожидание* рассогласования, которое обозначим a . Но об этом "истинном" значении рассогласования мы ничего не можем знать достоверно, остается о нем судить лишь с определенной **вероятностью** по ограниченному числу опытов. Поэтому наиболее полную оценку точности (вернее, погрешности) вычислительного эксперимента дает *доверительный интервал* (§ 5.5) для математического ожидания рассогласования. Так, например, может звучать вывод о точности в этом случае: с доверительной вероятностью 0,98 гарантируется погрешность угла атаки не более $0,3^\circ$. Критерием оценки точности тогда является соблюдение этой парой значений условий, приемлемых с точки зрения целей исследования.

Единственным практическим **недостатком** такой оценки может быть лишь необходимость знать закон распределения исследуемого рассогласования. Однако, во-первых, для оценки погрешности по подавляющему большинству параметров механических систем можно считать такое распределение нормальным, хотя бы приблизительно в некоторой области, а во-вторых, можно практиковать построение несимметричных доверительных интервалов, отражающих разную степень строгости требований по точности.

Доверительный интервал для оценки истинного значения рассогласования a по найденному значению выборочной средней величины рассогласования $\overline{\Delta u}$ при неизвестном значении σ , но известной несмещенной выборочной оценке среднего квадратического отклонения s , можно построить аналогично примеру § 5.5 на основании 5-ой строки табл. 10:

$$\overline{\Delta u} - t(\gamma, N - 1) \cdot \frac{s}{\sqrt{N}} < a < \overline{\Delta u} + t(\gamma, N - 1) \cdot \frac{s}{\sqrt{N}},$$

$$\text{где } \overline{\Delta u} = \frac{1}{N} \sum_{j=1}^r N_j \Delta u_j, \quad s^2 = \frac{1}{N-1} \sum_{j=1}^r N_j (\Delta u_j - \overline{\Delta u})^2,$$

$t(\gamma, N-1)$ определяется по распределению Стьюдента; N_j – число попаданий в j -й интервал (из r) наблюдаемых рассогласований Δu_j ; Δu_j – середина j -го интервала; N – общее число наблюдаемых значений Δu . Центр этого доверительного интервала определяется значением выборочной средней величины рассогласования $\overline{\Delta u}$. Размер доверительного интервала тем меньше, чем меньше доверительная вероятность γ , и чем больше число опытов N .

Естественно, при планировании вычислительного эксперимента следует стремиться к тому, чтобы такая оценка погрешности (т.е. доверительный интервал) не выходила за границы требуемой с точки зрения целей исследования погрешности $\pm\delta$, чего можно добиться разумным увеличением числа опытов N и уменьшением доверительной вероятности γ . Иными словами, следует стремиться к тому, чтобы доверительный интервал целиком укладывался внутри допустимой погрешности (например, от $-\delta$ до $+\delta$).

Если такого условия не удастся выполнить на данной серии опытов, то следует или увеличить число опытов N , или уменьшить доверительную вероятность γ . Однако последнее значительно слабее влияет на результат, тем более, что значения доверительной вероятности $\gamma < 0,7$ применять не желательно, так как это означает, что почти треть значений рассогласований может выходить за границы доверительного интервала (треть рассогласований принимает неконтролируемые значения).

Однако оценки точности с помощью доверительного интервала для математического ожидания рассогласования тоже недостаточно. Даже в том случае, когда рассеивание результатов мало, и гарантируется с определенной доверительной вероятностью, может существовать *систематическая погрешность*. Ее присутствие свидетельствует о **закономерности** рассогласования между оригиналом и моделью и не позволяет пользоваться ею.

Оценка систематической ошибки делается по величине выборочного среднего рассогласований, так как эта точечная оценка по своему смыслу характеризует среднее значение рассогласований, "присутствующее постоянно". Указанную оценку можно получить с помощью критерия Стьюдента, построенного на t -распределении из 5-ой строки табл. 10 в предположении идеального

случая: истинное значение математического ожидания погрешности равно нулю $a = 0$. По указанному критерию сравниваются две величины:

$$t = \frac{\overline{\Delta u}}{s} \sqrt{N} \quad \text{и} \quad t_{\text{крит.}}(1 - \alpha, N - 1),$$

где $t_{\text{крит.}}(1 - \alpha, N - 1)$ – предельное значение критерия, определяемое по таблице распределения Стьюдента при уровне значимости α (вероятности совершить ошибку первого рода: отвергнуть верную гипотезу об $a = 0$) с $N - 1$ степенями свободы.

Если $|t| < t_{\text{крит.}}$, то систематическая ошибка **незначима**, т.е. не существенна и можно считать $a = 0$. В случае противоположного неравенства: $|t| > t_{\text{крит.}}$ систематическая ошибка **значима**, т.е. не может считаться нулевой. В последнем случае математическая модель может считаться достаточно точной только при выполнении двух условий: доверительный интервал не выходит за границы требуемой с точки зрения целей исследования погрешности $\pm \delta$, а выборочное среднее погрешностей $\overline{\Delta u}$ пренебрежимо мало с точки зрения целей исследования, чтобы учитывать такую систематическую (закономерную) ошибку.

Непротиворечивость со статистической точки зрения может означать незначимость рассогласования, иными словами, неподверженность каким-либо закономерностям, непринципиальность – хаотичность. Последний термин и служит основой для построения критерия оценки непротиворечивости с помощью критерия согласия Пирсона χ^2 (§ 5.6). Для этого достаточно, чтобы рассогласование между моделью и оригиналом имело характер **простой ошибки измерений**, т.е. подчинялось нормальному закону распределения с нулевым математическим ожиданием $a = 0$.

По критерию Пирсона χ^2 для этого сравниваются две величины:

$$\chi_{\text{наблюдаемое}}^2 = \sum_{j=1}^r \frac{(N_j - Np_j)^2}{Np_j} \quad \text{и} \quad \chi_{\text{крит.}}^2(\alpha, n),$$

где p_j – вероятность попадания в j -й интервал (из r) нормально распределенной случайной величины с математическим ожиданием $a = 0$ и средним квадратическим отклонением $\sigma = s$, а $\chi_{\text{крит.}}^2(\alpha, n)$ – определяется по таблице распределения χ^2 при уровне значимости α (вероятности совершить ошибку первого рода: отвергнуть верную гипотезу о нормальном распределении рассогласования) с $n = r - 2$ степенями свободы.

Если $\chi_{\text{наблюдаемое}}^2 < \chi_{\text{крит.}}^2(\alpha, n)$, то различие статистического и гипотетического (нормального) законов распределения **незначимо**. Т.е. при заданном уровне значимости α гипотезу о поведении рассогласования эксперимента и "истины", как случайной ошибки измерений, можно принять и можно считать результаты вычислительного эксперимента не противоречащими реальному поведению оригинала. В случае противоположного неравенства:

$\chi^2_{\text{наблюдаемое}} > \chi^2_{\text{крит.}}(\alpha, n)$ расхождение **значимо** (закономерно, не может считаться случайным) и гипотезу следует отвергнуть, т.е. результаты вычислительного эксперимента противоречат реальному поведению оригинала.

Только в том случае, когда выполнены условия и требуемой точности, и непротиворечивости, можно считать результаты вычислительного эксперимента адекватными реальному поведению оригинала с доверительной вероятностью γ и уровнем значимости α в эксперименте из N опытов.

Таким образом, можно составить алгоритм проверки адекватности математической модели реальному поведению оригинала с помощью статистических критериев, предварительно задав допустимую погрешность $\pm\delta$, уровни значимости α_m , α и доверительную вероятность γ , исходя из целей исследования. В этом алгоритме строго соблюдается последовательность проверки статистических критериев, каждый следующий из которых опирается на вывод предыдущего.

1 ► Выбирается один из параметров объекта, для которого есть результаты наблюдения $\{U_i\}$ в N точках, и соответствующий параметр $\{u_i\}$, полученный в контрольном вычислительном эксперименте в тех же условиях в тех же точках.

2 ► Вычисляются разности $\Delta u_i = u_i - U_i$.

3 ► Вся область значений Δu разбивается на r интервалов таким образом, чтобы в каждый из них попало не менее пяти значений Δu_i .

4 ► Производится расчет количества попадания Δu_i в каждый j -й ($1 \leq j \leq r$) интервал – частот N_j .

5 ► Определяются статистические оценки параметров распределения случайной величины Δu : выборочное среднее $\overline{\Delta u} = \frac{1}{N} \sum_{j=1}^r N_j \Delta u_j$, где Δu_j – середина j -го интервала; и несмещенная оценка дисперсии $s^2 = \frac{1}{N-1} \sum_{j=1}^r N_j (\Delta u_j - \overline{\Delta u})^2$.

6 ► Для проверки **непротиворечивости**, т.е. подчиненности рассогласования нормальному закону распределения, применяется критерий согласия Пирсона

χ^2 . Вычисленное значение $\chi^2_{\text{набл.}} = \sum_{j=1}^r \frac{(N_j - Np_j)^2}{Np_j}$, где p_j – вероятность попадания

нормально распределенной случайной величины с математическим ожиданием $a = 0$ и средним квадратическим отклонением s в j -й интервал, сравнивается с табличным значением распределения Пирсона $\chi^2_{\text{крит.}}(\alpha; r - 2)$, где наиболее употребительные значения $\alpha = 0,05$ или $\alpha = 0,01$. Здесь число степеней свободы уменьшено на единицу, так как кроме выборочной средней используется найденная из выборки несмещенная оценка дисперсии. Если $\chi^2_{\text{набл.}} < \chi^2_{\text{крит.}}(\alpha; r - 2)$, то распределение Δu незначимо отличается от нормального, т.е. результаты вычислитель-

ного эксперимента можно считать не противоречащими реальному поведению оригинала. Если $\chi_{\text{набл}}^2 > \chi_{\text{крит}}^2(\alpha; r - 2)$, то значимое отличие распределения Δu от нормального свидетельствует о противоречии результатов вычислительного эксперимента реальному поведению оригинала и исследования адекватности следует прекратить.

Замечание. Проверка непротиворечивости (пригодности нормального закона распределения) проводится первой среди всех применяемых в данном алгоритме критериев и интервальных оценок, так как критерий согласия К. Пирсона χ^2 не требует сведений о законах распределения, а все последующие пункты **требуют знания закона** распределения и **основаны** на том, что Δu распределено по нормальному закону распределения с нулевым математическим ожиданием $a = 0$ и средним квадратическим отклонением s .

7► Для оценки систематической ошибки проверяется гипотеза о равенстве нулю математического ожидания ($a = 0$) рассогласования Δu с помощью критерия Фишера, для чего вычисляется статистика $t = \frac{\overline{\Delta u}}{s} \sqrt{N}$, которая сравнивается с $t(1 - \alpha_m; N - 1)$, определяемым по таблице распределения Стьюдента при уровне значимости α_m (0,05 или 0,01) и числе степеней свободы $N - 1$. Если $|t| > t(1 - \alpha_m; N - 1)$, то дальнейшие исследования адекватности нужно прекратить, так как это означает существование недопустимой систематической погрешности между результатами вычислительного эксперимента и реальным поведением оригинала. Если $|t| < t(1 - \alpha_m; N - 1)$, то систематическая погрешность отсутствует и можно продолжать исследования.

Замечание. Вывод об отсутствии систематической ошибки ($a = 0$) лишь **подтверждает** возможность исследования непротиворечивости в п. 6, а противоположный вывод – опровергает, т.е. делает его ничтожным.

8► Для оценки **точности** математической модели строится доверительный интервал для математического ожидания рассогласования при заданной доверительной вероятности γ (обычно 0,8; 0,9; 0,99; или 0,999):

$$\overline{\Delta u} - t(\gamma, N - 1) \cdot \frac{s}{\sqrt{N}} < a < \overline{\Delta u} + t(\gamma, N - 1) \cdot \frac{s}{\sqrt{N}},$$

где $t(\gamma, N - 1)$ определяется по таблице распределения Стьюдента. Если радиус доверительного интервала не превосходит допустимой погрешности

$$\left| t(\gamma, N - 1) \cdot \frac{s}{\sqrt{N}} \right| \leq \delta,$$

то математическую модель можно считать достаточно точной по отношению к оригиналу.

В дополнение к этому можно построить доверительный интервал для среднего квадратического отклонения рассогласования:

$$s \sqrt{\frac{N-1}{\chi^2(1-\gamma/2, N-2)}} < \sigma < s \sqrt{\frac{N-1}{\chi^2(\gamma/2, N-2)}}$$

используя распределение χ^2 .

9 ► Если по п. 7 можно считать математическую модель не противоречащей оригиналу, а по п. 8 и достаточно точной, то результаты расчетов адекватны реальному поведению оригинала.

Замечание. Если оценка точности математической модели оказывается во много раз лучше допустимой (иными словами, погрешность практически неразличима), то даже в отсутствии непротиворечивости математическую модель можно признать адекватной. Правда, получить с помощью предложенного алгоритма оценки систематической ошибки и точности можно только в том случае, если есть возможность загрузить требования непротиворечивости (т.е. задать существенно меньший уровень значимости α).

5.8. Основы статистического контроля качества технологических процессов

Рассмотренные в предыдущих параграфах статистические методы позволяют поставить любой технологический процесс под **математически** строгий контроль. Это означает, что, организовав специальным образом сбор статистического материала о параметрах производства, можно делать научно обоснованные (с оценкой ошибок) выводы о качестве процесса, тенденциях его изменения и о качестве продукции.

Регулирование качества технологических процессов на основе **текущего контроля** получило широкое развитие в 60-х годах XX века в основном в виде техники *контрольных карт*. Сбор и обработка статистической информации для них предельно формализованы, а преднамеренная ее фальсификация требует серьезной математической подготовки и недоступна на производстве. В последние годы стали появляться новые статистические методы контроля, учитывающие экономику и безопасность, что особенно важно в авиационной и космической технике.

В основе разработки контрольных карт лежат известные законы распределения. Так, по строкам 4 или 5 табл. 10 можно построить доверительный интервал как для математического ожидания по выборочному среднему, так и наоборот – для выборочного среднего по известному математическому ожиданию:

$$P(\bar{x} - \delta^+ < a < \bar{x} + \delta^-) = P(-\delta^+ < a - \bar{x} < \delta^-) = P(\delta^+ > \bar{x} - a > -\delta^-) = P(a + \delta^+ > \bar{x} > a - \delta^-) = \gamma.$$

Если в качестве математического ожидания a рассматривать **нормативное значение** контролируемого параметра технологического процесса (на которое он **должен** быть настроен), то за соблюдением этой настройки можно будет следить на основании замеров выборочной средней \bar{x} . Тогда, исходя из заданной доверительной вероятности γ и величин допуска в меньшую δ^- и в

большую δ^+ сторону можно будет отмечать случаи нарушения заданных требований к технологическому процессу.

Если закон распределения не удастся подобрать даже приближенно, то используется неравенство Чебышева:

$$P(|\bar{x} - a| < \delta) \geq 1 - \frac{\sigma^2}{\delta^2},$$

правда, в этом случае необходимо знание дисперсии σ^2 , которую можно оценить дополнительными **предварительными** исследованиями процесса. Это же может понадобиться и при традиционном построении доверительных интервалов.

Методы оценки σ основываются на теории точечных оценок (§ 5.3) и предлагают в качестве σ использовать одно из следующих значений:

$$\frac{s}{\tilde{c}_m}, \quad \frac{\bar{s}}{c_2^*} = \frac{1}{c_2^* k} \cdot \sum_{i=1}^k s_i, \quad \frac{R}{d_2} = \frac{x_{\max} - x_{\min}}{d_2},$$

где \tilde{c}_m, c_2^*, d_2 – поправки на смещение [33].

Собственно **контрольная карта** представляет собой график изменения контролируемого параметра по времени, на котором заранее нанесены *контрольные границы (границы допуска)* в соответствии с границами доверительного интервала. На рис. 49 показана для примера контрольная карта средних \bar{x} . Контроль заключается в нанесении статистических данных на карту и принятии решения о вмешательстве в технологический процесс. Процесс считается *статистически подконтрольным*, пока значения \bar{x} лежат внутри контрольных границ. При выходе \bar{x} за эти пределы производство следует приостановить для перенастройки или переналадки оборудования.

Контрольная карта зазора. Номинал: 5, max: 7, min: 2.

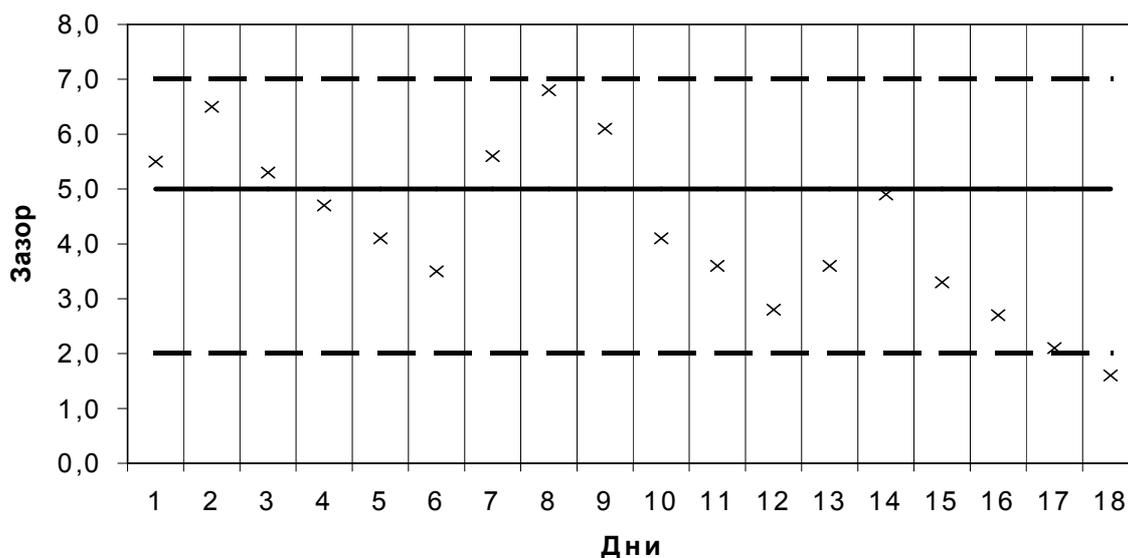


Рис. 49.

Подготовительная работа требует для определения границ допуска знания не только σ , но и объемов выборок для контроля N_i , и доверительной вероятности γ (или уровня значимости – вероятности ошибки $\alpha = 1 - \gamma$). Кроме того, если используется гипотеза о каком-либо определенном законе распределения, то ее необходимо проверить по критерию Пирсона χ^2 (§ 5.6).

Для объемов контрольных выборок N_i следует назначать постоянно одно и то же небольшое нечетное число N , например, 5 или 7, что позволит упростить расчеты среднего и определение медианы и размаха.

В качестве доверительной вероятности на практике принимаются "удобные" для применяемого закона распределения числа, например, для нормального закона: 0,9973 (из правила 3σ), 0,99 или 0,95, а для неравенства Чебышева 0,89 (из правила 3σ).

Приведем перечень распространенных видов контрольных карт по контролируемым параметрам:

1) карта средних значений " \bar{x} ";

2) карта индивидуальных значений " x_i " – при объеме выборки $N = 1$ решение принимается после выхода за границы 4 – 6 последовательных значений;

3) карта медиан " \tilde{x} " – основывается на приблизительно нормальном распределении медианы с математическим ожиданием a и дисперсией $\sigma^2 \frac{\pi}{2N}$;

4) карта " \bar{x}/s ", являющаяся комбинацией из двух карт: для \bar{x} и для s , имеющей контрольные границы

$$0 < s < \left(1 + \frac{u_\alpha}{c_2^*} \sqrt{1 - c_2^{*2}} \right) \bar{s}, \text{ где } \bar{s} = \frac{1}{k} \sum_{i=1}^k s_i;$$

5) карта " \bar{x}/R ", являющаяся комбинацией из двух карт: для \bar{x} и для R ,

6) карта " \tilde{x}/R ", являющаяся комбинацией из двух карт: для \tilde{x} и для R ,

7) карта "p" – среднего процента брака с контрольными границами

$$p - u_\alpha \sqrt{\frac{p(1-p)}{N}} < \frac{x}{N} < p + u_\alpha \sqrt{\frac{p(1-p)}{N}},$$

в этом случае N необходимо брать достаточно большим,

8) карта "Np" – модификация карты p,

9) карта "c" – дефектов, отказов, распределенных по закону Пуассона, с контрольными границами

$$c - u_\alpha \sqrt{c} < x < c + u_\alpha \sqrt{c},$$

где c – параметр распределения Пуассона,

10) карта "u" – обобщение карты "c" на процент дефектов, отказов.

Карты 1 – 6 называются *контрольными картами по измеримым признакам*, карты 7 – 10 – по *неизмеримым*.

Техника контрольных карт не всегда позволяет обеспечить приемлемый уровень текущего контроля. Это особенно ярко проявляется в сложных наукоемких и дорогостоящих технологических процессах, где практически невоз-

можно отследить все существенные параметры. Кроме того, и сам процесс контроля становится дорогостоящим и высокотехнологичным, если строго следовать технологии контрольных карт.

Поэтому усилия ученых в последние годы сосредоточились на разработке таких статистических методов, которые позволяли бы по минимуму выходных параметров, определяющих потребительские свойства продукции, оптимально с экономической точки зрения осуществлять и контроль, и управление сложным технологическим процессом.

Отечественные ученые разработали ряд методов *контроля и управления случайными процессами* [2], основной целью которых было обеспечить **экономичное** обслуживание по состоянию. Следует особо отметить, что все эти математически строгие результаты доведены до алгоритмов, пригодных к практическому использованию. Здесь излагаются два из этих методов, сформулированные в виде теорем.

Один из них пригоден для управления техническим состоянием сложной системы с монотонным показателем ее работоспособности. Рассматривается $\xi(t)$ – случайный **монотонно неубывающий** во времени t процесс, о развитии которого можно судить по некоторому контролируемому признаку, наблюдаемое значение которого в момент времени t_k обозначено x_k . Это соответствует обычной технологической практике, когда контрольные мероприятия проводятся через определенные промежутки времени, не обязательно равные. Суммарные затраты на наблюдение и регулировку системы в период ее работоспособности, т.е. пока $x_k < L$ (L – пороговое значение, соответствующее отказу) составляют сумму C , а штраф (потери при отказе, т.е. выходе системы за порог отказа, когда $x_k > L$) – величину A .

Если для такого процесса **изменение** признака $\Delta x_k = x_k - x_{k-1}$ на очередном интервале времени контроля не зависит от текущего времени, т.е. не зависит от предыстории, и распределено по известному (хотя бы эмпирически) закону: $P(\Delta x_k < z) = F(z)$, то минимум удельных затрат на контроль и обслуживание достигается при прекращении эксплуатации (для ремонта, замены и т.п.) в момент времени t_{k-1}^* , когда впервые будет **нарушено** неравенство:

$$1 - P(\Delta x_k < L - x_{k-1}) \leq \frac{C}{A(k-1)},$$

или, что то же самое, когда впервые **выполнится** неравенство:

$$F(L - x_k) < 1 - \frac{C}{Ak}.$$

Алгоритм действий при организации управления качеством по такому методу рассмотрим на численном примере, проиллюстрированном на рис. 50. На левом графике приведен эмпирический закон распределения $F(z)$, полученный предварительным сбором статистической информации, по которому с помощью *единичного жребия* (§ 3.5) можно симулировать разовое изменение контролируемого признака Δx_k . На правом графике представлено монотонное развитие процесса от состояния полностью исправной системы (признак $x = 0$) к отказу (признак $x = L$), характерное для большинства реальных функциональных

систем авиационной и космической техники. Видно, как с течением времени контролируемый признак ("дефектность", количество отказавших подсистем и т.п.) стремится к порогу L , за которым система считается отказавшей. Если, как это обычно бывает, затраты на наблюдение и регулировку в процессе нормальной эксплуатации C меньше штрафа при отказе A , то момент времени t^* прекращения эксплуатации настанет до того, как контролируемый признак достигнет порогового значения. А это именно та стратегия, которая необходима при эксплуатации по состоянию.

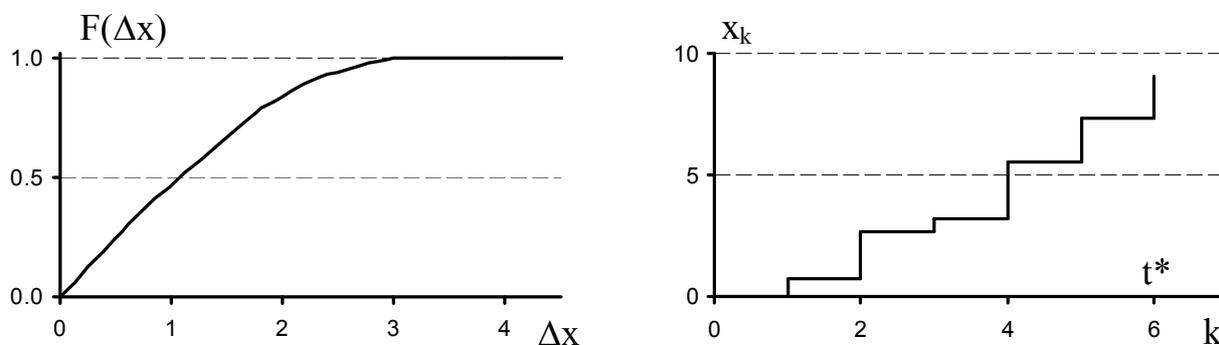


Рис. 50.

Графики рис. 50 построены по данным расчетной табл. 12 с помощью последовательности случайных чисел (во втором столбце) при заданных значениях: $L = 10$, $C/A = 0,5$. Момент времени t^* наступает в нашем примере при $k = 6$, когда второе из приведенных в теореме условий впервые выполняется (см. последние два столбца табл. 12), но пороговое значение контролируемого показателя еще не вышло за предел $L = 10$, т.е. отказ еще не наступил.

Таблица 12.

| k | $F(z)$ | Δx_k | x_k | $L - x_k$ | $F(L - x_k)$ | $1 - \frac{C}{Ak}$ |
|-----|--------|--------------|--------|-----------|--------------|--------------------|
| 1 | 0,3721 | 0,7626 | 0,7626 | 9,2374 | 1,0000 | 0,5000 |
| 2 | 0,8195 | 1,9209 | 2,6835 | 7,3165 | 1,0000 | 0,7500 |
| 3 | 0,2455 | 0,4960 | 3,1795 | 6,8205 | 1,0000 | 0,8333 |
| 4 | 0,9169 | 2,3206 | 5,5002 | 4,4998 | 1,0000 | 0,8750 |
| 5 | 0,7939 | 1,8343 | 7,3345 | 2,6655 | 0,9718 | 0,9000 |
| 6 | 0,7697 | 1,7568 | 9,0913 | 0,9087 | 0,4389 | 0,9167 |

Ступенчатость графика изменения x_k по k отображает дискретность профилактических работ, когда на следующий межконтрольный период предполагается постоянство показателя. На самом деле он может расти постоянно и опасность выхода за предельное значение присутствует сразу после k -го момента (с этого момента остаток до отказа $L - x_k$ может быть "перекрыт" за один межконтрольный период – см. табл. 12), поэтому именно в этот момент времени систему необходимо снять с эксплуатации и произвести мероприятия по ее обновлению: замену, ремонт, переналадке, настройке и т.п. Однако следует заметить, что приведенный алгоритм имеет математически строгое доказательство, т.е. он гарантирует недостижение отказной ситуации в реальных условиях при соблюдении условий применимости.

Второй рассматриваемый здесь метод весьма перспективен для учета изменения интенсивности отказов, когда она растет у **стареющей** системы. Пусть интенсивность отказов не постоянна $\lambda \neq \text{const}$, $\lambda'(t) \geq 0$, т.е. увеличивается со временем по известному (хотя бы эмпирическому) закону: $\lambda = \lambda(t)$. Обозначим:

T_1 – среднее время замены исправной системы,

T_2 – среднее время замены неисправной системы (обычно на практике $T_2 > T_1$ из-за затрат на ликвидацию последствий отказа),

T_3 – момент времени **предупредительной замены** системы (пока еще исправной),

$F(\omega) = 1 - e^{-\int_0^{\omega} \lambda(t) dt}$ – вероятность безотказной работы системы в момент времени ω ,

$p(t, \Delta t)$ – вероятность того, что в момент времени t система исправна и проработает еще в течение времени Δt ,

$p(t, \Delta t) \Rightarrow \max$ – критерий оптимальности на **всем сроке** жизни системы с учетом многократного ее восстановления. Тогда оптимальное T_3 определится из нелинейного уравнения:

$$\frac{T_1}{T_2 + \Delta t} = 1 - \frac{1}{1 - F(T_3) + \lambda(T_3) \int_0^{T_3} [1 - F(\omega)] d\omega}.$$

Приемочный контроль, в отличие от текущего, осуществляется изготовителем (выходной контроль) или потребителем (приемка) по готовой продукции для отбраковки и не имеет непосредственной обратной связи с производством. Если объем продукции велик или она сложна, возникает необходимость по результатам проверки ее **части** сделать заключение о ее пригодности. Это позволяет сделать *выборочный метод*, разработанный именно для такого рода задач и предназначенный для получения выводов о свойствах генеральной совокупности по свойствам выборки.

Основой выборочного метода является правильный отбор (§ 5.2) информации. При этом для приемочного контроля используются как однократные, так и многократные выборки. Рассмотрим применение однократных выборок, для которых необходимо определить объем N и приемочное число c – норму браковки. Т.е. для принятия всей партии из M изделий будем считать необходимым, чтобы в проверяемой выборке объема N число бракованных изделий было меньше c .

Для определения N и c необходимо вспомнить, что при $N < M$ возможны два вида ошибок (когда выборка неверно отражает качество всей партии):

– ошибка I рода, когда по результатам выборочного контроля **бракуется годная** партия, ее вероятность обозначим α ,

– ошибка II рода, когда по результатам выборочного контроля **негодная** партия **принимается**, ее вероятность обозначим β .

Поэтому вместо одного критического значения p^* **доли брака** приходится назначать два: p_α – *допускаемый уровень качества*, и p_β – *недопустимый уровень качества*. Между ними ($p_\alpha < p_\beta$) находится область неопределенности, которую необходимо исключить при организации приемочного контроля. Именно для замены этих двух границ вычисляется приемочное число c .

Из теории вероятностей известно, что если доля брака определяется вероятностью p , то вероятность обнаружения среди N изделий, отобранных из M , ровно k бракованных вычисляется по формуле:

$$P_N(k, p) = \frac{C_{pM}^k C_{M-pM}^{N-k}}{C_M^N} = \frac{\binom{pM}{k} \binom{M-pM}{N-k}}{\binom{M}{N}}.$$

Это *гипергеометрическое распределение* при больших M и малых p аппроксимируется *распределением Пуассона*:

$$P(k, p) = \frac{(Np)^k e^{-Np}}{k!}.$$

Тогда вероятность принятия партии с **допускаемым** уровнем качества p_α , т.е. случая, когда в выборке качества p_α число бракованных изделий не превосходит приемочного числа c , определится суммой:

$$p_\alpha = \sum_{k=0}^c P_N(k, p_\alpha) = \sum_{k=0}^c \frac{C_{p_\alpha M}^k C_{M-p_\alpha M}^{N-k}}{C_M^N} = \sum_{k=0}^c \frac{\binom{p_\alpha M}{k} \binom{M-p_\alpha M}{N-k}}{\binom{M}{N}} = 1 - \alpha,$$

а вероятность принятия партии с **недопустимым** уровнем качества p_β , т.е. случая, когда в выборке качества p_β число бракованных изделий не превосходит приемочного числа c , определится суммой:

$$p_\beta = \sum_{k=0}^c P_N(k, p_\beta) = \sum_{k=0}^c \frac{C_{p_\beta M}^k C_{M-p_\beta M}^{N-k}}{C_M^N} = \sum_{k=0}^c \frac{\binom{p_\beta M}{k} \binom{M-p_\beta M}{N-k}}{\binom{M}{N}} = \beta.$$

В этих двух уравнениях присутствуют два неизвестных N и c , которые из них и определяются. Таким образом, строится однократная выборка для приемочного контроля качества.

При многократных выборках появляется возможность построить процедуру *последовательного анализа* для принятия решения, когда на каждом i -м шаге по результатам очередной i -й выборки можно сделать три вывода:

- а) в случае числа бракованных изделий $k < \bar{c}_i$ принять партию,
- б) в случае числа бракованных изделий $k > \bar{\bar{c}}_i$ забраковать партию,
- в) в случае $\bar{c}_i < k < \bar{\bar{c}}_i$ продолжить контроль.

Глава 6. Основы многомерного статистического анализа

6.1. Классификация задач многомерного статистического анализа

В предыдущей главе рассматривались методы первичной обработки информации и приемы анализа единственной случайной величины – этот материал иногда называют одномерным статистическим анализом. В этом смысле установление статистических закономерностей, как цель статистического анализа, завершалось решением небольшого круга задач, связанных с оценкой вероятности тех или иных предположений – гипотез о поведении единичных случайных величин. Однако на практике значительно чаще встречаются объекты с двумя и более случайными величинами, и при разработке моделей сложных систем необходимость многомерного статистического анализа становится очевидной. В таких моделях приходится изучать не только характерные особенности отдельных случайных *факторов*, но и их **взаимодействие**. При этом возможны различные подходы к выявлению и оценке такого взаимодействия.

Объект исследования можно представить в виде "черного ящика" с "входами" и "выходами" (рис. 51), среди которых различают:

$\mathbf{X} = (x_1, x_2, \dots, x_k)$ – вектор **входных контролируемых** переменных, которыми **можно управлять** в исследовании;

$\mathbf{Z} = (z_1, z_2, \dots, z_p)$ – вектор **входных контролируемых** переменных, которыми **невозможно управлять** в исследовании;

$\mathbf{E} = (e_1, e_2, \dots, e_f)$ – вектор **входных неконтролируемых и неуправляемых** переменных (*шум*);

$\mathbf{Y} = (y_1, y_2, \dots, y_g)$ – вектор **выходных** переменных.



Рис. 51.

Переменные \mathbf{X} , \mathbf{Z} и \mathbf{Y} называются факторами. Если фактор принимает фиксированные, детерминированные значения, то они называются уровнями фактора. Факторы \mathbf{Z} могут быть случайными. Переменные \mathbf{E} тоже случайные, хотя могут и не описываться законами распределения (могут быть не стохастического, не вероятностного вида), но происхождение их не является предметом исследования, они проистекают из-за погрешностей эксперимента или моделирования. Если факторы \mathbf{Z} случайны, то и выходные переменные \mathbf{Y} необходимо

рассматривать как случайные. В частности, такого рода случайности могут проистекать из-за наложения шума E .

Рассмотрим некоторые виды задач многомерного статистического анализа, которые могут возникнуть при изучении сложных *многофакторных систем*.

1) Существует ли **связь** между отдельными факторами (любыми из: u_i, u_k, x_s, z_t, z_v).

2) Если между какими-то факторами есть связь, то насколько она **тесная**.

3) Если между какими-то факторами есть связь, то какой функцией ее можно **представить**.

4) Какие входные факторы оказывают на определенные выходные наибольшее **влияние**.

5) Какие входные факторы можно **отбросить** из процесса изучения на основании их слабого, сравнимого с шумом, влияния.

6) Существуют ли **неучтенные** факторы, которые необходимо рассматривать ввиду их существенного влияния на выходные.

7) Существуют ли **обобщенные** факторы, которыми можно заменить несколько рассматриваемых.

8) Как связаны между собой **зашумленные** факторы.

9) Каковы **характеристики шума**.

10) Как **выделить** "полезную" информацию из зашумленной.

Все эти задачи можно решить с помощью методов многомерного статистического анализа, включающего в себя:

- корреляционный анализ,
- регрессионный анализ,
- конфлюэнтный анализ,
- дисперсионный анализ,
- факторный анализ,
- теорию фильтрации.

Каждый из этих разделов – совокупность методов и приемов математической статистики. Исторически так сложилось, что четкого разграничения указанных разделов и методов между разделами не существует. Выбор методов диктуется лишь конкретной практической задачей. Так, например, задачи 1, 2 вышеприведенного списка решаются методами корреляционного анализа, задача типа 3 – регрессионного анализа. Задачи 4, 5, 6 относятся к задачам дисперсионного анализа, 7 – факторного, а 8, 9 – конфлюэнтного. Теория фильтрации позволяет решить задачи типа 10 и частично 9.

Каждый из перечисленных видов статистического анализа, по сути, предназначен для обоснования той или иной *статистической математической модели* изучаемого оригинала. Корреляционный анализ позволяет получить корреляционную модель, регрессионный анализ – регрессионную модель, конфлюэнтный – конфлюэнтную, дисперсионный – дисперсионную. Их виды и назначение будут рассмотрены в последующих параграфах.

Возможность решения перечисленных задач вытекает из физического **смысла** основных **числовых характеристик** случайных величин: дисперсии или среднего квадратического отклонения и коэффициента корреляции. Какими методами можно получить их точечные оценки по выборочным наблюдениям, было рассмотрено в предыдущих параграфах. Сосредоточим здесь внимание на их осмыслении.

Поскольку речь идет о многомерном статистическом анализе, рассмотрим подробнее простейшую систему из **двух случайных величин** ξ и η , принимающих возможные дискретные значения x_i и y_j . Дисперсии этих величин:

$$\sigma_{\xi}^2 = \sum_i (x_i - a)^2 p_i, \quad \sigma_{\eta}^2 = \sum_j (y_j - b)^2 p_j,$$

где a и b соответствующие математические ожидания, характеризуют **рассеяние** возможных значений, а **ковариация**, являющаяся тоже моментом второго порядка,

$$\text{cov}(\xi, \eta) = \sum_{i,j} (x_i - a)(y_j - b)p_{ij}$$

не дает характеристику рассеяния, поскольку может принимать отрицательные значения. Из этой формулы видно, что ковариация положительна, если значения первой случайной величины, большие a , **чаще** появляются **вместе** со значениями второй случайной величины, большими b (a меньшие – с меньшими). В противоположном случае ковариация отрицательна. Т.е. на графике связи (но не зависимости!) этих величин в первом случае можно наблюдать **тенденцию** возрастания (рис. 52а), а во втором – убывания (рис. 52б). Эта тенденция может быть больше или меньше "размыта" в зависимости от согласованности отклонений каждой из величин от своего математического ожидания (рис. 52а и 52г). Таким образом, ковариация показывает, насколько связь между двумя случайными величинами близка к строгой линейной, но и одинаково отмечает и слишком большую **случайность**, и слишком большую **нелинейность** этой связи.

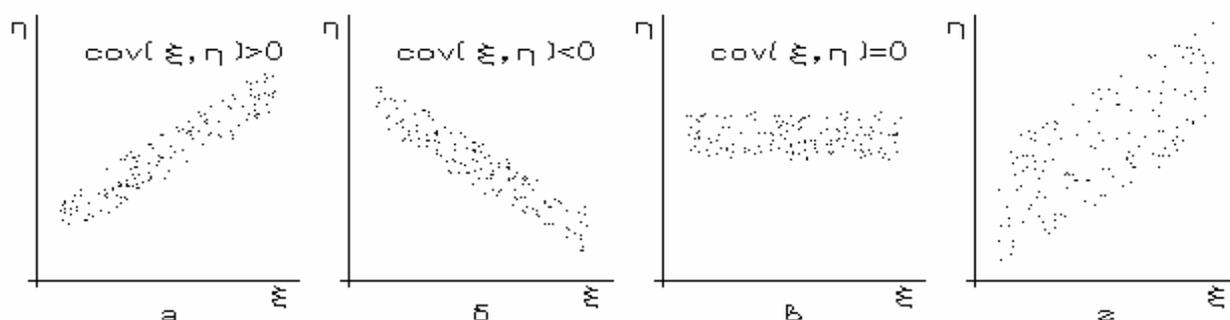


Рис. 52.

С другой стороны, если ξ и η независимы, то согласно известной теореме умножения вероятностей $p_{ij} = p_i p_j$ и ковариация обращается в **нуль** (рис. 52в):

$$\text{cov}(\xi, \eta) = \sum_{i,j} (x_i - a)(y_j - b)p_{ij} = \sum_i (x_i - a)p_i \sum_j (y_j - b)p_j = E(\xi - a)E(\eta - b) = 0.$$

Обратное утверждение не имеет места: из равенства нулю ковариации независимость случайных величин в общем случае **не следует**. (Однако можно показать, что для системы **нормально** распределенных случайных величин понятия коррелированности и зависимости тождественны.)

Вычислим дисперсию линейной комбинации двух случайных величин:

$$\begin{aligned}\sigma_{\alpha\xi+\beta\eta}^2 &= \sum_{i,j} [(\alpha x_i + \beta y_j) - (\alpha a + \beta b)]^2 p_{ij} = \sum_{i,j} [\alpha(x_i - a) + \beta(y_j - b)]^2 p_{ij} = \\ &= \alpha^2 \sum_i (x_i - a)^2 p_i + \beta^2 \sum_j (y_j - b)^2 p_j + 2\alpha\beta \sum_{i,j} (x_i - a)(y_j - b)p_{ij} = \\ &= \alpha^2 \sigma_\xi^2 + \beta^2 \sigma_\eta^2 + 2\alpha\beta \cdot \text{cov}(\xi, \eta)\end{aligned}$$

Как видно, она тоже является линейной комбинацией исходных дисперсий с **квадратами** коэффициентов, но с **добавлением** члена, зависящего от ковариации. Если последняя равна нулю, то дисперсия линейной комбинации случайных величин есть линейная комбинация дисперсий с квадратами коэффициентов. Этот замечательный факт позволил дать определение таким *случайным величинам*, ковариация которых равна нулю – **некоррелированные**. В противном случае случайные величины называются **коррелированными**, т.е. между ними наблюдается **связь**. Это не означает, однако, что они взаимно зависимы. Если случайные величины независимы, то они обязательно некоррелированы, но если они зависимы, то могут быть как коррелированными, так и некоррелированными.

Подробное рассмотрение понятия корреляции необходимо для четкого представления свойств многомерного закона распределения: его выражение должно включать в себя не только математические ожидания составляющих a и b и их дисперсии σ_ξ и σ_η , но и *коэффициент корреляции* между составляющими $\rho_{\xi\eta}$. Конечной целью и основной **задачей корреляционного анализа** и является построение совместного **закона** распределения системы случайных величин.

Задачу регрессионного анализа составляет исследование вида и формы корреляции: какой функцией можно **представить** связь случайных величин ξ и η (принимающих, соответственно, значения x и y) – *регрессию* – уравнение для условной средней, вычисленной при каждом x : $\bar{y}_x = f(x)$. (Условной средней \bar{y}_x называют среднее арифметическое значение величины η при определенном значении ξ .) При этом несущественно, случайна или детерминирована величина, принимаемая за аргумент.

Некоторые авторы считают регрессионный анализ частью корреляционного, некоторые считают его методы самостоятельными. Однако эти методы могут предложить лишь то или иное выражение связи, они не отвечают на вопрос о наличии собственно функциональной **зависимости** – такая постановка вопроса возможна только в профессиональном плане исследования объекта, а не в формальном математическом.

Задачей конфлюэнтного анализа является изучение **структуры** случайных величин, находящихся в некотором взаимодействии.

Наибольшее развитие у классика многомерного статистического анализа Р. Фишера получил дисперсионный анализ, основной задачей которого является **сравнение дисперсий** разных случайных величин или различных способов вычисления дисперсий.

Факторный анализ решает задачу поиска минимального числа обобщенных факторов, заменяющих исходное множество.

Задачей теории фильтрации является выделение исходного сигнала из искаженной или неполной информации.

6.2. Понятие о корреляционном анализе

Корреляционный анализ – группа статистических методов установления формы и тесноты корреляционной связи между факторами. Методы корреляционного анализа будем рассматривать, исходя из решаемых с его помощью практических задач, перечисленных в предыдущем параграфе.

Связь двух факторов можно выявить по величине коэффициента корреляции $\rho_{\xi\eta}$, для которого нам надо получить статистическую оценку r_{xy} . Простейший прием для этого – применение метода моментов, т.е. вычисление выборочного коэффициента корреляции как смешанного момента второго порядка – см. § 5.3. Такая оценка, однако, является смещенной, поэтому при $N < 30$ ею следует пользоваться только как предварительной. Значения коэффициента корреляции позволяют различать почти **некоррелированные** величины, когда он приблизительно равен нулю, и почти **линейную связь**, когда он близок к единице.

Множество факторов (более двух), влияющих на состояние системы, требует использования множества парных коэффициентов корреляции, поэтому может быть описано корреляционной таблицей или матрицей (симметричной) коэффициентов корреляции.

Для проверки гипотезы о **нормальном** распределении совокупности двух случайных величин ξ и η **первым** шагом по результатам **специальных дополнительных** наблюдений необходимо получить выборочные оценки S_x , S_y и r_{xy} параметров распределения σ_ξ , σ_η и $\rho_{\xi\eta}$. Далее необходимо проверить частные их распределения, что можно сделать по известному критерию Пирсона χ^2 (§ 5.6). Если они не нормальны, то можно попытаться найти такие их функции, которые были бы нормально распределены. В случае неудачи такого поиска исследования корреляции надо прекратить ввиду невозможности описать эту систему двумерным нормальным законом.

Если частные распределения выдержали проверку на нормальный закон, то вторым шагом надо проверить коррелированность ξ и η . Для этого можно использовать функцию:

$$\chi_2^2 = \frac{1}{1 - \rho_{\xi\eta}^2} \left[\left(\frac{\bar{x} - a}{\sigma_\xi} \right)^2 - 2\rho_{\xi\eta} \frac{\bar{x} - a}{\sigma_\xi} \frac{\bar{y} - b}{\sigma_\eta} + \left(\frac{\bar{y} - b}{\sigma_\eta} \right)^2 \right],$$

имеющую распределение χ_2^2 с двумя степенями свободы, **если** ξ и η нормально коррелированы. Эту выборочную функцию можно включить дополнительной строкой в табл. 10. После

этого по данным N **следующих дополнительных** наблюдений вычисляют N значений χ_2^2 . Проверку распределения этих значений можно делать визуально, так как:

$$p(\chi_2^2) = \frac{1}{2} e^{-\frac{1}{2}\chi_2^2}, \quad P(\chi_2^2) = 1 - e^{-\frac{1}{2}\chi_2^2} \quad \text{и} \quad \lg(1 - P(\chi_2^2)) \approx -0,217\chi_2^2.$$

Таким образом, достаточно на полулогарифмической бумаге нанести точки с координатами $\{(\chi_2^2)_i; (N - i + 1/2)/N\}$, где значения $(\chi_2^2)_i$ пронумерованы по возрастанию, и проверить случайность расположения их вокруг прямой, проходящей через точку $\{0; 1\}$ с наклоном $-0,217$.

При более полном анализе на третьем шаге необходимо построить доверительный интервал для коэффициента корреляции и проверить гипотезу о равенстве его интересующему нас в задаче значению. Эти процедуры можно проделать уже известным способом, используя закон распределения. Однако на практике при этом не обойтись без компьютера, так как распределение выборочного коэффициента корреляции записывается весьма громоздким выражением через гамма-функцию (строка 15 табл. 10). В важном частном случае при $N > 30$ и $\rho_{\xi\eta} < 0,5$ распределение выборочного коэффициента корреляции близко к нормальному с математическим ожи-

данием $\rho_{\xi\eta}$ и средним квадратическим отклонением $\frac{1 \pm \rho_{\xi\eta}^2}{\sqrt{N-1}}$.

В частном случае равенства нулю истинного значения коэффициента корреляции $\rho_{\xi\eta} = 0$ распределение упрощается и дает возможность, так же как и в случае нормальности исследуемых величин, использовать две строки (14 и 15) табл. 10. Таким образом, можно сформулировать **критерий** для проверки гипотезы о **коррелированности** нормальных слу-

чайных величин. Если вычисленное значение $t = \frac{r_{xy} \sqrt{N-2}}{\sqrt{1-r_{xy}^2}}$ окажется больше, чем крити-

ческое $t_{1-\alpha}(N-2)$, определенное из таблиц распределения Стьюдента при заданном уровне значимости α и числе степеней свободы $N-2$, то имеющиеся данные не подтверждают гипотезу о коррелированности.

В результате проведения корреляционного анализа определяется выборочная оценка r_{xy} для основной числовой характеристики связи в законе распределения системы двух случайных величин – коэффициента корреляции $\rho_{\xi\eta}$. Вместе с выборочными оценками \bar{x} , \bar{y} , s_x , s_y остальных числовых характеристик a , b , σ_ξ , σ_η это позволяет явным образом записать предполагаемый закон распределения вероятностей для (ξ, η) . Запись гипотетического закона распределения составляет так называемую **корреляционную модель** – один из видов статистических математических моделей. Так, например, **нормальный** закон распределения **системы** двух случайных величин выглядит следующим образом:

$$f(\xi, \eta) = \frac{1}{2\pi\sigma_\xi\sigma_\eta\sqrt{1-\rho_{\xi\eta}^2}} e^{-\frac{1}{2(1-\rho_{\xi\eta}^2)} \left[\frac{(\xi-a)^2}{\sigma_\xi^2} - \frac{2\rho_{\xi\eta}(\xi-a)(\eta-b)}{\sigma_\xi\sigma_\eta} + \frac{(\eta-b)^2}{\sigma_\eta^2} \right]}.$$

Тесноту связи между факторами можно оценивать доверительным интервалом для коэффициента корреляции. Но из-за сложности вычислений

функции распределения коэффициента корреляции более распространенным является ее оценка по корреляционному отношению.

Изучим полную дисперсию случайной величины η , представляемой нами в качестве функции от ξ . Ее следует рассматривать, как оценку рассеяния относительно математического ожидания b , и можно преобразовать следующим образом:

$$\begin{aligned}\sigma_{\eta}^2 &\equiv D(\eta) \equiv E(\eta - b)^2 = E(\eta - \bar{y}_x + \bar{y}_x - b)^2 = \\ &= E(\eta - \bar{y}_x)^2 + E(\bar{y}_x - b)^2 + 2E(\eta - \bar{y}_x)(\bar{y}_x - b).\end{aligned}$$

Последнее слагаемое равно нулю, так как:

$$E(\eta - \bar{y}_x)(\bar{y}_x - b) = (\bar{y}_x - b)E(\eta - \bar{y}_x) = 0,$$

поэтому можно ввести следующие обозначения:

$$\sigma_{\eta}^2 = E(\eta - \bar{y}_x)^2 + E(\bar{y}_x - b)^2 \equiv \sigma_{\eta/x}^2 + \delta_{\eta/x}^2.$$

По выражению второго слагаемого $\delta_{\eta/x}^2$ видно, что оно показывает **дисперсию линии регрессии** $\bar{y}_x = f(x)$ (широту размаха) **относительно математического ожидания b** , т.е. измеряет степень **влияния фактора ξ** на фактор η . Первое слагаемое $\sigma_{\eta/x}^2$ дает **дисперсию случайной величины η** (разброс точек) **относительно линии регрессии \bar{y}_x** и измеряет **влияние неучтенных факторов** на η . Именно им ($\sigma_{\eta/x}^2$) можно пользоваться, как мерой тесноты связи случайных величин: чем меньше эта дисперсия, тем теснее связь (см. рис. 52а и 52г).

Однако большее распространение для этой меры получило корреляционное отношение:

$$\Theta_{\eta/x}^2 \equiv \frac{\delta_{\eta/x}^2}{\sigma_{\eta}^2} = 1 - \frac{\sigma_{\eta/x}^2}{\sigma_{\eta}^2},$$

которое обращается в **единицу при отсутствии** влияния неучтенных факторов ($\sigma_{\eta/x}^2 = 0$), т.е. в случае однозначной функциональной связи, представляемой линией регрессии. В нуль оно обращается тогда и только тогда, когда $\delta_{\eta/x}^2 = 0$ и влияние фактора ξ на фактор η не прослеживается, т.е. при $\bar{y}_x = b + \text{const}$, иными словами, при линейной регрессии с нулевым углом наклона.

Эти свойства наводят на мысль о том, что корреляционным отношением можно пользоваться так же, как коэффициентом корреляции. Это действительно так, более того, в случае **линейной** регрессии последний является частным случаем первого и оба могут оценивать тесноту связи.

При практических исследованиях необходимо учитывать смещенность статистических оценок коэффициента корреляции и корреляционного отношения. Это смещение происходит в сторону **завышения** тесноты связи тем больше, чем меньше объем выборки.

Таким образом, с учетом основного правила статистического анализа (каждое утверждение должно быть проверено) алгоритм проведения корреляционного анализа выглядит следующим образом.

- 1► Получение статистических оценок \bar{x} , \bar{y} , s_x , s_y , r_{xy} точечных характеристик a , b , σ_ξ , σ_η , $\rho_{\xi\eta}$ случайных величин.
- 2► Проверка гипотез о законах распределения изолированных случайных величин ξ и η .
- 3► Проверка гипотезы о коррелированности случайных величин ξ и η .
- 4► Проверка гипотезы о равенстве коэффициента корреляции определенному числу $\rho_{\xi\eta} = \rho_0$.
- 5► Оценка тесноты связи случайных величин ξ и η .
- 6► Построение доверительного интервала для коэффициента корреляции $\rho_{\xi\eta}$.
- 7► Построение совместного закона распределения системы случайных величин ξ и η .

6.3. Дисперсионный анализ

Дисперсионный анализ разработан Р. Фишером в 1920 году и представляет собой группу методов математической статистики для анализа результатов наблюдений, зависящих от **нескольких** одновременно действующих факторов, как случайных, так и детерминированных, как наблюдаемых, так и ненаблюдаемых.

Идея дисперсионного анализа заключается в **разбиении** общей дисперсии изучаемой случайной величины на независимые составляющие, каждая из которых характеризует влияние того или иного фактора или их взаимодействие. Последующее **сравнение** этих дисперсий позволяет оценить **значимость влияния** факторов на исследуемую величину.

При разработке математического аппарата Фишеру потребовалось сделать следующие предположения дисперсионного анализа, необходимые для обоснованности выкладок и выводов.

1) Исследуемые факторы стохастически **независимы**. С точки зрения способов отбора информации это означает независимость выборочных результатов наблюдения (отдельных выборок или слоев – они не преобразуются друг в друга с помощью какого-либо алгоритма).

2) Исследуемые факторы, каждый по отдельности, подчиняются **нормальным** законам распределения.

3) **Дисперсии** σ_i^2 исследуемых факторов **однородны** (априори приблизительно одного порядка).

Возможности и приемы ослабления этих предположений изучаются в конце данного параграфа.

Рассмотрим подробно, как производится разбиение общей дисперсии и какой вид и смысл имеют составляющие. Прделаем это сначала для простейшего случая **единственного** входного фактора, оказывающего предположительное **влияние** на выходной фактор системы. Входной фактор в дальнейших рассуждениях явно не присутствует, хотя результаты наблюдений "охватывают" возможные его изменения. (В качестве примера конкретной задачи такого вида можно привлечь пример § 7.4 и рис. 54).

Для представления результатов *однофакторного эксперимента* воспользуемся идеей конглоэнтного анализа о структурировании выходного фактора:

$$y_{ji} = b_j + \varepsilon_{ji},$$

где $j = 1, 2, \dots, k$ – номер уровня (значения) исследуемого входного фактора, $i = 1, 2, \dots, N_j$ – порядковый номер замера, y_{ji} – результат единичного замера выходного фактора η системы, b_j – математическое ожидание (генеральное среднее) фактора η при **j-м уровне** исследуемого входного фактора, ε_{ji} – независимые стохастические компоненты наблюдений, распределенные по единому нормальному закону с нулевым математическим ожиданием и дисперсией σ^2 . Один исследуемый фактор сам по себе независим, таким образом, все требования дисперсионного анализа выполнены.

Однако для построения каких-либо вычислений одной идеи конглоэнтного анализа о структурировании выходного фактора недостаточно. Значения b_j не известны и известными быть не могут, поэтому возможен только **анализ** этой структуры. Если мы исследуем предположительное влияние входного фактора на выходной, то при каждом уровне входного фактора результаты замеров выходного должны группироваться вокруг разных значений b_j . Эту **гипотезу** о различии математических ожиданий b_j результатов измерений выходного фактора при разных уровнях (j) входного фактора нам и надо проверить. Иначе говоря, b_j можно рассматривать как функцию от номера j уровня входного фактора. Поэтому представим эту величину математического ожидания, изменяющуюся в зависимости от исследуемого входного фактора, в виде суммы:

$$b_j = \mu + T_j,$$

где μ – математическое ожидание фактора η при **всех** уровнях исследуемого входного фактора, T_j – добавок к μ от **влияния** исследуемого входного фактора. Тогда можно написать математическую модель для **однофакторного дисперсионного анализа** (дисперсионную модель) исследуемого влияния в виде:

$$y_{ji} = \mu + T_j + \varepsilon_{ji},$$

позволяющую проводить некоторые исследования.

Заметим, что за неимением значений b_j их можно заменить оценками \bar{y}_j и представить структурное уравнение в виде:

$$y_{ji} = \bar{y}_j + \delta_{ji},$$

где δ_{ji} – независимые стохастические компоненты наблюдений, тоже распределенные по единому нормальному закону с нулевым математическим ожиданием и дисперсией σ^2 . О независимости здесь следует говорить лишь в смысле независимости результатов наблюдений при каждом уровне исследуемого входного фактора. Такое структурирование дает возможность разбить сумму квадратов отклонений, используемую для статистической оценки общей дисперсии, на части:

$$\begin{aligned} s^2 &= \sum_{j=1}^k \sum_{i=1}^{N_j} (y_{ji} - \bar{y})^2 = \sum_{j=1}^k \sum_{i=1}^{N_j} (y_{ji} - \bar{y}_j + \bar{y}_j - \bar{y})^2 = \\ &= \sum_{j=1}^k \sum_{i=1}^{N_j} (y_{ji} - \bar{y}_j)^2 + \sum_{j=1}^k \sum_{i=1}^{N_j} (\bar{y}_j - \bar{y})^2 + 2 \sum_{j=1}^k \sum_{i=1}^{N_j} (y_{ji} - \bar{y}_j)(\bar{y}_j - \bar{y}) = \\ &= \sum_{j=1}^k \sum_{i=1}^{N_j} (y_{ji} - \bar{y}_j)^2 + \sum_{j=1}^k N_j (\bar{y}_j - \bar{y})^2 + 2 \sum_{j=1}^k (\bar{y}_j - \bar{y}) \sum_{i=1}^{N_j} (y_{ji} - \bar{y}_j) \end{aligned}$$

Последнее слагаемое правой части обращается в **нуль**, так как внутренняя сумма нулевая по определению \bar{y}_j . Первое слагаемое дает оценку рассеяния **внутри** серий наблюдений – при каждом уровне j исследуемого входного фактора – и позволяет вычислить оценку остаточной (внутренней) дисперсии, т.е. влияния всех **неучтенных** факторов:

$$s_0^2 = \frac{1}{N - k} \sum_{j=1}^k \sum_{i=1}^{N_j} (y_{ji} - \bar{y}_j)^2.$$

Второе слагаемое дает оценку рассеяния **между** сериями наблюдений – при различных значениях исследуемого входного фактора – и позволяет вычислить межгрупповую дисперсию, т.е. **влияние** изменения исследуемого входного **фактора**:

$$s_A^2 = \frac{1}{k - 1} \sum_{j=1}^k N_j (\bar{y}_j - \bar{y})^2.$$

(Можно образно представить s_A^2 как размах "плавного отслеживания" изменения выходного фактора в результате изменения исследуемого входного фактора, а s_0^2 как разброс индивидуальных результатов наблюдения при "замороженном" уровне исследуемого фактора.)

Таким образом, получено основное уравнение дисперсионного анализа, которое может быть записано в двух видах:

$$\sum_{j=1}^k \sum_{i=1}^{N_j} (y_{ji} - \bar{y})^2 = \sum_{j=1}^k \sum_{i=1}^{N_j} (y_{ji} - \bar{y}_j)^2 + \sum_{j=1}^k N_j (\bar{y}_j - \bar{y})^2,$$

$$\text{или } (N - 1) \cdot s^2 = (N - k) \cdot s_0^2 + (k - 1) \cdot s_A^2.$$

Из последней записи видно, что при совпадении s_A^2 и s_0^2 они дают одновременно и оценку s^2 общей дисперсии. Этот результат позволяет сделать замечательный вывод: если все выборочные данные подчиняются **одному** и тому же нормаль-

ному закону распределения (с общими математическим ожиданием и дисперсией), то различие между s_A^2 и s_0^2 должно быть **незначимо**. Для проверки этой гипотезы можно построить критерий с использованием закона распределения выборочной функции из 13-й строки табл. 10 § 5.4, называемый критерием Фишера. Тогда из попадания отношения $\frac{s_A^2}{s_0^2}$ в критическую область следует необходимость отверг-

нуть гипотезу о едином нормальном законе распределения (с общими математическим ожиданием и дисперсией). В этом случае **влияние** исследуемого входного фактора следует считать **значимым** – с ним нельзя не считаться, так как изменение выходного фактора при изменении входного **неслучайно** – **закономерно**.

Однако возникает вопрос, как выбирать критическую область для критерия Фишера. Очевидно, что значимое превосходство s_A^2 над s_0^2 (т.е. существенное, а не простое превышение) означает существенно большее влияние исследуемого входного фактора, чем прочих неучтенных. Поэтому критическая область должна быть односторонней при выбранном уровне значимости α :

$$\frac{s_A^2}{s_0^2} > F_{1-\alpha}(k-1, N-k),$$

где $F_{1-\alpha}(k-1, N-k)$ определяется по таблицам распределения Фишера при вероятности $1 - \alpha$ и двух числах степеней свободы: $k - 1$ для **большой** дисперсии (в данном случае s_A^2) и $N - k$ для **меньшей** (s_0^2).

Что касается некритической области, то ее следует рассмотреть подробнее в силу зависимости распределения Фишера от двух чисел степеней свободы. Если в противоположность первому случаю межгрупповая дисперсия **значимо меньше** остаточной:

$$\frac{s_0^2}{s_A^2} > F_{1-\alpha}(N-k, k-1),$$

(заметим, что здесь числа степеней свободы поменялись местами, так как первое соответствует большей дисперсии) то влияние исследуемого входного фактора несущественно и им можно пренебречь, так как при этом влияние изменения исследуемого входного фактора на выходной "забывается" влиянием остальных неучтенных факторов.

В случае, когда нельзя говорить о значимом превосходстве одной из дисперсий над другой, влияние исследуемого входного фактора **сравнимо** с погрешностью эксперимента или влиянием неучтенных факторов, поэтому конкретный вывод невозможен.

С одной стороны, здесь открывается поле деятельности для *последовательного анализа*. С другой стороны, это тот самый случай, когда можно гово-

речь о незначимости различия между s_A^2 и s_0^2 , близости s_A^2 , s_0^2 и оценки s^2 общей дисперсии, т.е. о единстве закона распределения (с общими математическим ожиданием и дисперсией) для различных уровней исследуемого фактора.

На рис. 53 показаны различные случаи экспериментальных данных, характеризующихся различными соотношениями дисперсий. По оси абсцисс отложены лишь **номера** уровней исследуемого входного фактора, но не его физическая величина – так обычно строятся исследования в дисперсионном анализе, чтобы не привносить лишней информации. Основываясь лишь на зрительном восприятии этого рисунка, нельзя сказать, есть ли зависимость функции, отложенной по ординате, от параметра, отложенного по абсциссе. Этого нельзя сказать даже в том случае, если расположить очередность уровней исследуемого входного фактора в порядке возрастания частных средних, соответствующих этим уровням, которые на рисунке обозначены кружочками и соединены сплошной линией. Несмотря на это дисперсионный анализ позволяет сделать достаточно уверенный вывод о влиянии исследуемого входного фактора на выходной.

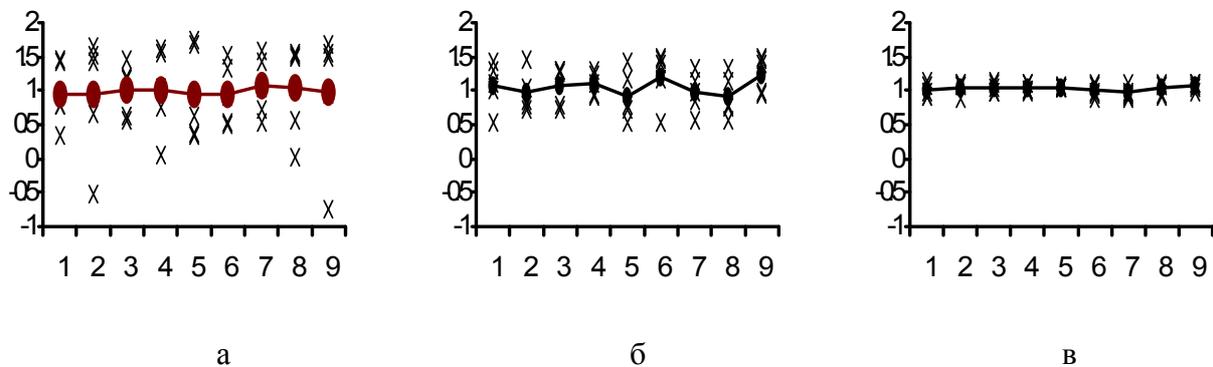


Рис. 53.

В случае "а" большая дисперсия – остаточная (внутренняя):

$$\frac{s_0^2}{s_A^2} = 8,07 > F_{1-\alpha}(N - k, k - 1) = 5,15, \text{ что свидетельствует о значительном влиянии неуч-}$$

тенных факторов, которые "забивают" возможную зависимость от исследуемого входного фактора. В этих условиях естественно считать эту зависимость несущественной. В случае "б"

больше уже межгрупповая дисперсия, но отношение дисперсий не достигает критического значения по критерию Фишера: $\frac{s_A^2}{s_0^2} = 1,21 < F_{1-\alpha}(k - 1, N - k) = 3,04$, следовательно, сле-

дательный вывод о влиянии или невлинии исследуемого входного фактора нельзя. В случае "в" межгрупповая дисперсия не только больше, но и **значимо** больше остаточной:

$$\frac{s_A^2}{s_0^2} = 9,02 > F_{1-\alpha}(k - 1, N - k) = 3,04, \text{ поэтому необходимо сделать вывод о существенности влияния исследуемого входного фактора.}$$

До сих пор в данном параграфе рассматривался *однофакторный дисперсионный анализ*. Его целью было оправдание или опровержение предполагаемой дисперсионной модели, выделяющей единственный фактор. Однако практика чаще ставит задачи, в которых необходимо рассматривать системы с несколькими входными факторами, определяющими выходной регистрируемый

фактор. Для таких задач проводится *многофакторный дисперсионный анализ*, изучающий *многофакторную дисперсионную модель*:

$$y_{ij\dots m} = \mu + T_i + S_j + \dots + Q_m + \varepsilon_{ij\dots m},$$

где $y_{ij\dots m}$ – результат эксперимента, в котором фактор T принял i -ый уровень, фактор S – j -ый, фактор Q – m -ый уровень. Вышеприведенный математический аппарат принципиально не изменяется, лишь основное уравнение дисперсионного анализа принимает более громоздкий вид. Критерий Фишера для попарного сравнения дисперсий применяется и в многомерном случае, но для сравнения **нескольких** дисперсий (нескольких факторов) существуют и другие критерии, которые можно найти в специальной литературе.

Рассмотрим подготовку к проведению дисперсионного анализа с точки зрения обеспечения необходимых требований его математического аппарата.

Независимость исследуемых факторов является важнейшим условием дисперсионного анализа, пренебрегать которым нельзя из-за опасности получить бессмысленные или неверные выводы. В случае априорной неопределенности в этом плане необходимо провести *корреляционный анализ* на базе отдельного **специально** поставленного эксперимента. Тогда можно будет опираться хотя бы на некоррелированность факторов (для коррелированных факторов проводить дисперсионный анализ бессмысленно).

Если исследуемые факторы оказались зависимыми или коррелированными, то необходимо попытаться подобрать другие факторы или выбрать только независимые. Большую помощь в этом могут оказать *метод главных компонент* и *факторный анализ* (§ 8.1).

В случае неподчинения исследуемых факторов **нормальному закону** распределения используется следующий прием. Можно предположить, что этот "недостаток" наблюдаемых величин обусловлен их зависимостью от каких-то других **скрытых** факторов, которые распределены нормально. Тогда всю выборку следует **перегруппировать** в новые слои, а затем провести анализ по всем дисперсиям. Для того чтобы такие поиски не велись вслепую, можно привлечь метод *главных компонент* или *факторный анализ* (§ 8.1).

Однородность дисперсий в слоях при различных значениях исследуемых факторов можно проверить по критерию Фишера, согласно 12-й строке табл. 10 § 5.4. Если они не однородны, можно перейти к новому фактору:

$$g(x) = q \cdot \int \frac{dx}{h(x)}.$$

При этом функция $h(x)$ выбирается из условия описания связи

математического ожидания a и среднего квадратического отклонения σ : $\sigma = h(a)$, а коэффициент $q = h(a) \cdot g'(a)$.

Таким образом, вырисовывается следующая последовательность действий при дисперсионном анализе по выборочным результатам:

1 ► Проверка независимости (или некоррелированности) исследуемых факторов методами корреляционного анализа. При необходимости обеспечение некоррелированности с помощью замены факторов.

2► Проверка нормального распределения исследуемых факторов по критерию согласия Пирсона. При необходимости замена факторов.

3► Проверка однородности дисперсий по критерию Фишера. При необходимости замена факторов.

4► Разбиение общей дисперсии на составляющие в соответствии с задачей исследований.

5► Вычисление необходимых межгрупповых и остаточных дисперсий и проверка гипотез о значимости их различия с помощью критерия Фишера.

(6) ► Анализ отклонений средних от общего среднего (проверка гипотезы о равенстве математических ожиданий), при этом используется критерий знаков для k величин: $\frac{\bar{y}_i - \bar{\bar{y}}}{s_0} \sqrt{N_i}$, а при больших N_i и k еще и проверка нор-

мального распределения k величин (4-я или 5-я строка табл. 10 § 5.4):

$$\frac{\bar{y}_i - b}{\sigma} \sqrt{N_i} \text{ или } \frac{\bar{y}_i - b}{s} \sqrt{N_i}.$$

(7) ► Если гипотеза о равенстве математических ожиданий отвергнута, то можно определить доверительные интервалы для них с помощью распределения Стьюдента с $N - k$ степенями свободы для функции $\frac{\bar{y}_i - b_i}{s_0} \sqrt{N_i}$.

Последние два пункта не являются обязательными для собственно дисперсионного анализа, а служат для завершения общей картины связи выборочных оценок точечных параметров законов распределения слоев (групп) выборки.

Некоторые примеры применения многомерного статистического анализа помещены в главе 7. Достаточно полное описание разнообразных классических примеров можно найти в [29].

6.4. Регрессионный анализ

Регрессионный анализ предназначен для решения единственной задачи: получения теоретического уравнения регрессии $\eta(\xi) = f(\xi, \lambda)$, вид которого задается, исходя **из особенностей** изучаемой системы случайных величин, а параметры λ определяются по выборочным данным. Случайная величина η рассматривается как функция от неслучайной величины ξ . Однако, детерминированность (неслучайность) ξ не оказывает влияния на проведение и результаты регрессионного анализа. Регрессия – функциональная зависимость, *аппроксимирующая* (заменяющая) статистическую зависимость средних значений рассматриваемых факторов (переменных) $\bar{y}_x = f(x)$.

В общем случае регрессия, получаемая на основе регрессионного анализа с целью дальнейшего использования функциональной зависимости вектора выходных факторов от входных в математической модели исследуемого явления, носит название регрессионной модели. В качестве таковых рассматриваются соотношения вида:

$$Y = \Phi(\mathbf{X}, \mathbf{Z}, \Lambda) + E,$$

где Λ – вектор коэффициентов регрессионной модели, подлежащих определению из эксперимента, а вид вектор-функции Φ предполагается выбранным заранее на основании известных свойств объекта исследования.

Наиболее общим из распространенных видов регрессионных моделей является линейная по базисным функциям $\mathbf{F}(\mathbf{X}, \mathbf{Z})$ от входных факторов, имеющая общий вид, использующий скалярное произведение: $y = (\Lambda, \mathbf{F}(\mathbf{X}, \mathbf{Z})) + e$, где одна выходная переменная y представлена скалярным произведением, с аддитивной (суммируемой) помехой e . В этом случае коэффициенты регрессионной модели λ_i умножаются на некоторые **ортогональные** базисные функции $f_i(\mathbf{X}, \mathbf{Z})$, вид которых определяется свойствами объекта исследований. Это могут быть:

– тригонометрические функции $f_{(2n-1)i}(\mathbf{X}) = \cos(nx_i)$; $f_{2ni}(\mathbf{X}) = \sin(nx_i)$,

если явление имеет признаки периодической ограниченной величины;

– системы ортогональных полиномов (многочленов Эрмита, Лежандра, Лагерра и т.п.);

– "полиномиальные переменные" $f_i(\mathbf{X}) = x_i; \dots$; $f_{n+i}(\mathbf{X}) = x_i^2; \dots$;

$f_{2n+i}(\mathbf{X}) = x_{i-1}x_i; \dots$;

т.е. регрессионная модель может быть представлена в виде конечных сумм соответствующих рядов, отражающих все их свойства (периодичность, ограниченность, наличие или отсутствие экстремумов и точек перегиба).

Простейшей регрессионной моделью является линейная непосредственно по входным факторам функция:

$$y = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k + e.$$

В общем случае **полиномиальной** регрессии (с помощью многочленов) необходимо помнить, что эксперимент с k уровнями по некоторому входному фактору сможет зафиксировать изменения выходного фактора, описываемые полиномом не более $k - 1$ порядка. Иными словами, для линейной регрессии достаточно иметь 2 уровня исследуемого фактора, так как для определения 2 коэффициентов линейной функции необходимо иметь 2 условия (через две точки проходит единственная прямая). Однако этот крайний случай не для регрессии – она строится через облако экспериментальных точек, соответствующих необходимому количеству уровней входного фактора.

Рассмотрим последовательные шаги решения общей задачи регрессионного анализа.

1 ► **Задание вида** линии *регрессии* – важнейший этап. Основные соображения о виде функции $f(\xi, \lambda)$ должны диктоваться только реальными **физическими** свойствами изучаемого явления, например, возрастание/убывание, выпуклость/вогнутость, наличие замечательных точек (известные значения координат отдельных точек, экстремумов и перегибов), асимптот, периодов, свойств симметрии, гладкости и т.д. И только дополнительными мотивами могут слу-

жить графические построения выборочной корреляционной зависимости. В результате этого этапа следует получить аналитическую запись функции $f(\xi, \lambda)$ с неизвестными параметрами λ , которые предстоит определить.

2► Вычисление *выборочных оценок* параметров предполагаемого теоретического закона распределения: $\bar{x}, \bar{y}, s_x, s_y, r_{xy}$.

3► *Проверка гипотезы* о равенстве нулю коэффициента корреляции (§ 6.2). Этот шаг нужен для выяснения вопроса о необходимости дальнейшего анализа. Если исследуемые случайные величины по результатам выборки можно считать некоррелированными, то бесполезно пытаться строить регрессию и следует прекратить анализ или вернуться к пункту 1.

4► *Проверка гипотез* о законах распределения исследуемых случайных величин ξ и η , как изолированных. Это необходимо для следующих шагов, в частности, для метода наибольшего правдоподобия, и может быть проделано с помощью критерия согласия Пирсона χ^2 (§ 5.6).

5► Для отыскания *параметров линии регрессии* λ используется *метод наибольшего правдоподобия* (§ 5.3), дающий несмещенные их оценки λ^* и не предъявляющий каких-либо требований к законам распределения случайных величин. Правда, применить этот метод можно только тогда, когда последние **известны**. Поэтому и необходим предыдущий шаг.

(6)► Уравнение регрессии можно считать окончательным только после подтверждения дисперсионным анализом (§ 6.3) **незначимости** отличия s_R^2 от s_0^2 . Здесь $s_R^2 = \frac{1}{k-v} \sum_{i=1}^k N_i [\bar{y}_i - f(x_i; \lambda)]^2$ характеризует рассеяние выборочных средних **вокруг линии регрессии**, т.е. "погрешность" замены статистической регрессии (линии, связывающей средние по слоям \bar{y}_i) аппроксимирующей зависимостью $y = f(x, \lambda)$; s_0^2 – определенное в § 5.4 **остаточное** рассеяние – характеризует разброс отдельных экспериментальных точек вокруг средних по слоям, т.е. "погрешность" самого эксперимента; v – число параметров функции регрессии, определенных из выборки. Проверка производится по критерию Фишера для $\frac{s_R^2}{s_0^2}$ с $(k-v, N-k)$ степенями свободы (дополнительная строка в

табл. 10 § 5.4). **Незначимость** отличия s_R^2 от s_0^2 характеризует такую "малость" отклонений вокруг линии регрессии, которую можно "объяснить" погрешностью самого эксперимента.

(7)► Построение доверительных интервалов для оценок параметров распределения λ вокруг их оценок λ^* , а также самой случайной величины η ("коридор" вокруг эмпирической линии регрессии) в общем случае представляет собой весьма сложную задачу. В этой ситуации необходимо из законов распределения ξ и η выразить законы распределения параметров λ , а это не всегда

возможно. Поэтому чаще всего прибегают к приближенным законам и приближенным оценкам.

(8) ► Для частного случая установления **линейной** функции регрессии $\eta = \alpha_0 + \alpha_1(\xi - a)$ между **нормально** распределенными случайными величинами аппарат регрессионного анализа разработан достаточно полно – до аналитического вида или до использования распространенных таблиц (см. табл. 10 § 5.4). Но этот аппарат накладывает **дополнительные требования** на дисперсию значений η при каждом x : она должна быть или постоянна, или пропорциональна известной функции от x . **Проверка гипотезы о линейности регрессии** выполняется с помощью дисперсионного отношения $\frac{s_R^2}{s_0^2}$, где знаменатель определен

в § 5.4, а $s_R^2 = \frac{1}{k-2} \sum_{i=1}^k N_i (\bar{y}_i - Y_i)^2$ соответствует рассеянию средних \bar{y}_i вокруг линии регрессии $Y_i = a_0 + a_1(x_i - \bar{x})$. Число степеней свободы у s_R^2 равно $k-2$, так как помимо выборочного среднего \bar{y} (которым определяется a_0) предполагается вычисленным и коэффициент линейности a_1 . Кстати, именно s_R^2 может служить оценкой общей дисперсии σ теоретической регрессии. Если по распределению Фишера с $(k-2, N-k)$ степенями свободы s_R^2 значительно превышает s_0^2 , то гипотеза о линейности должна быть отвергнута.

(9) ► В случае подтверждения гипотезы о линейности регрессии критерии согласия и доверительные интервалы для параметров теоретической регрессии $\eta = \alpha_0 + \alpha_1(\xi - a)$ можно построить, исходя из распределения Стьюдента с $N-2$ степенями свободы. Ему в этом случае подчиняются следующие выборочные функции:

$$\frac{a_0 - \alpha_0}{s_{a0}}, \quad \frac{a_1 - \alpha_1}{s_{a1}}, \quad \frac{Y - \eta}{s_y},$$

$$\text{где } s_{a0} = \frac{s_R}{\sqrt{N}}, \quad s_{a1} = \frac{s_R}{\sqrt{\sum N_i (x_i - \bar{x})^2}}, \quad s_y = s_R \cdot \sqrt{\frac{1}{N} + \frac{(x - \bar{x})^2}{\sum N_i (x_i - \bar{x})^2}}.$$

Доверительные интервалы для коэффициентов α_0 и α_1 однозначны. Для η он зависит от x (см. последнюю формулу): чем дальше x от \bar{x} , тем шире границы доверительного интервала $|Y - \eta|$.

Метод наибольшего правдоподобия приводит к задаче оптимизации типа нелинейного (в общем случае) программирования (§ 4.5). В **частном** случае нормально распределенных величин он сводится к *методу наименьших квадратов*, разработанному в XIX веке Лежандром и Гауссом. Вообще говоря, применять метод наименьших квадратов можно и в отсутствие нормального распределения – формально, однако оценки параметров в этом случае будут иметь смещение тем большее, чем более асимметрично истинное распределение ис-

пользуемых выборочных функций. Поэтому к полученным таким "формальным" способом *регрессионным моделям* следует относиться с большой осторожностью.

Метод наименьших квадратов основан на отыскании таких значений параметров λ функциональной зависимости заданного вида $f(\xi, \lambda)$, которые минимизируют величину суммы квадратов отклонений вычисленных значений функции от соответствующих N пар наблюдаемых значений (x_i, y_i) , см. рис. 54:

$$J(\lambda) = \sum_{i=1}^N [y_i - f(x_i, \lambda)]^2 \Rightarrow \min .$$

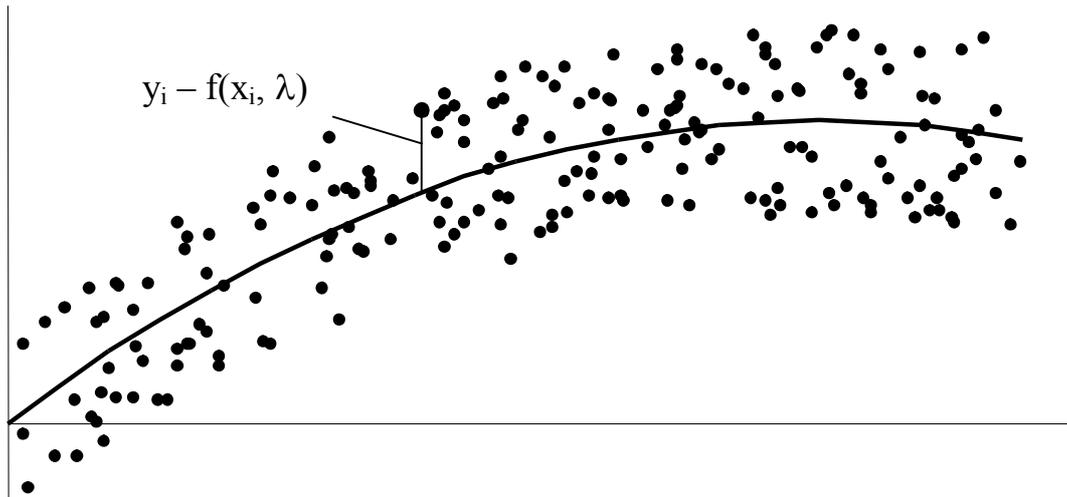


Рис. 54.

Следовательно, параметры λ_j функции $f(x, \lambda)$ определяются из условия минимума функции нескольких аргументов $J(\lambda)$, т.е. из системы $m + 1$ уравнений:

$$\frac{\partial J}{\partial \lambda_j} = -2 \sum_{i=1}^N [y_i - f(x_i, \lambda)] \cdot \frac{\partial f}{\partial \lambda_j} = 0, \quad j = 0, 1, 2, \dots, m,$$

в которой неизвестными являются все $m + 1$ параметров λ_j (столько же, сколько и уравнений), а все значения x_i и y_i известны из эксперимента.

В важном частном случае полиномиальной регрессии (многочленами) вида:

$$y = f(x, \lambda) = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m = \sum_{j=0}^m a_j x^j$$

в результате несложных преобразований получается так называемая "система нормальных уравнений метода наименьших квадратов":

$$\sum_{j=0}^m a_j \cdot \sum_{i=1}^N x_i^{j+k} = \sum_{i=1}^N y_i x_i^k, \quad k = 0, 1, 2, \dots, m,$$

которая в развернутом виде выглядит следующим образом:

$$\left\{ \begin{array}{l} a_0 \cdot N + a_1 \cdot \sum_{i=1}^N x_i + a_2 \cdot \sum_{i=1}^N x_i^2 + \dots + a_m \cdot \sum_{i=1}^N x_i^m = \sum_{i=1}^N y_i, \\ a_0 \cdot \sum_{i=1}^N x_i + a_1 \cdot \sum_{i=1}^N x_i^2 + a_2 \cdot \sum_{i=1}^N x_i^3 + \dots + a_m \cdot \sum_{i=1}^N x_i^{m+1} = \sum_{i=1}^N y_i x_i, \\ + a_0 \cdot \sum_{i=1}^N x_i^m + a_1 \cdot \sum_{i=1}^N x_i^{m+1} + a_2 \cdot \sum_{i=1}^N x_i^{m+2} + \dots + a_m \cdot \sum_{i=1}^N x_i^{2m} = \sum_{i=1}^N y_i x_i^m. \end{array} \right.$$

Таким образом, задача отыскания регрессии, например, для экспериментальной зависимости полиномом m -ой степени сводится к решению системы неоднородных **линейных алгебраических уравнений** с $m + 1$ неизвестными коэффициентами полинома.

Испытания надежности некоторой аппаратуры дали результаты, сформированные в корреляционную таблицу 13, в которой x задает время наладки (испытания, доработки), y – время безотказной эксплуатации конкретного прибора, а в основной части таблицы – число опытов, закончившихся с результатами x , y (время задается в %% от нормы).

Таблица 13.

| y | x | | | | |
|-----|----|----|-----|-----|-----|
| | 60 | 80 | 100 | 120 | 140 |
| 60 | 6 | 3 | | | |
| 80 | 2 | 16 | 16 | 2 | 2 |
| 100 | 2 | 23 | 49 | 19 | 4 |
| 120 | | 5 | 8 | 13 | 4 |
| 140 | | 1 | 4 | 6 | 15 |

Предположим для начала, что исследуемая зависимость может быть представлена линейной регрессией $f(x, a) = a_0 + a_1 x$. Тогда система нормальных уравнений метода наименьших квадратов приобретает вид:

$$\left\{ \begin{array}{l} a_0 \cdot N + a_1 \cdot \sum_{i=1}^N x_i = \sum_{i=1}^N y_i, \\ a_0 \cdot \sum_{i=1}^N x_i + a_1 \cdot \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i x_i. \end{array} \right.$$

Эту систему линейных алгебраических уравнений можно решить любым из известных методов (§ 4.1) после вычисления коэффициентов. В данном примере:

$$N = 200, \quad \sum_{i=1}^N x_i = 20440, \quad \sum_{i=1}^N x_i^2 = 2179200, \quad \sum_{i=1}^N y_i = 20520, \quad \sum_{i=1}^N y_i x_i = 2148000,$$

что дает $a_0 = 44,998$, $a_1 = 0,56362$. Результат в этом случае годен **лишь для оценок в пределах исследованного диапазона**, ибо из всех существенных физических свойств величины времени безотказной работы отражает лишь возрастание.

Из теории надежности известно, что существует некоторый предельный уровень надежности, т.е. величина y должна стремиться к "насыщению" и переставать зависеть от дальнейшего увеличения x . Поэтому для выявления такого рода тенденции линейная регрессия неприменима. Следует искать другой вид зависимости – не полиномиальной, так как многочлены не имеют асимптот.

Можно составить уравнения метода наименьших квадратов, вообще говоря, для любого вида зависимости, однако не всегда эти уравнения просто решаются. Поэтому пойдем в своем анализе другим путем: исследуем другое важное физическое свойство рассматриваемого

мой величины – выпуклость – с помощью многочлена второй степени, имеющего это свойство: $f(x, \mathbf{a}) = a_0 + a_1x + a_2x^2$.

Решение системы из 3 нормальных уравнений метода наименьших квадратов для 3 коэффициентов a_0, a_1, a_2 в этом случае можно получить с помощью, например, правила Крамера через определители, для чего понадобятся дополнительные суммы:

$$\sum_{i=1}^N x_i^3 = 241456000, \quad \sum_{i=1}^N x_i^4 = 27694100000, \quad \sum_{i=1}^N y_i x_i^2 = 234176000.$$

В результате решения системы уравнений получим: $a_0 = 49,902, a_1 = 0,46503, a_2 = 0,00047465$, что свидетельствует о наличии в зависимости выпуклости вниз. Это противоречие с физическим смыслом кривой надежности говорит о существенной недостаточности статистического материала для отражения явления "насыщения". Поэтому попытки подобрать более сложный вид регрессии на данном статистическом материале скорее всего окончатся безрезультатно. Необходимо или сделать новую выборку, или расширить данную за счет дополнительных результатов наблюдения.

Замечание. Наличие в системе линейных алгебраических уравнений коэффициентов, различающихся на несколько порядков, существенно осложняет решение с приемлемой точностью. Поэтому следует применять масштабирование исходных параметров. В данном примере удобен один масштаб по x и y : 100, который приводит параметры к величинам, **близким** к 1. В более общем случае для полиномиальных регрессий бывает полезно масштабирование со

"сдвигом":
$$x = \frac{x - x_0}{m_x}.$$

В более общем случае **нелинейной** регрессии бывает полезно привести исходную зависимость с помощью каких-либо замен переменных к виду, близкому к линейному или квадратичному, и уже для этой новой зависимости искать регрессию.

Необходимо отметить, что наблюдаемая зависимость может быть и сугубо неоднозначной **корреляционной** – такой, как это показано на рис. 54, и более привычной, **функциональной**, однозначной по смыслу, как это имеет место, например, для немногочисленных экспериментальных данных. Тогда, например, при построении поляры самолета, отыскание регрессии принимает характер так называемого "*сглаживания*" экспериментальной зависимости. Вообще говоря, такая замена сложной функциональной зависимости более простой – например, аналитической – носит название *аппроксимации* (см. § 4.1). Так что отыскание регрессии, сглаживание и аппроксимация – эквивалентные названия одной и той же задачи. Во всех этих случаях с помощью метода наибольшего правдоподобия можно отыскать функцию **удобного** вида $y = f(x, \lambda)$, **заменяющую** наблюдаемую с приемлемой точностью. Точность аппроксимации можно оценить по величине:

$$\sum_{i=1}^N [y_i - f(x_i, \lambda)]^2 \quad \text{или} \quad \sum_{i=1}^N |y_i - f(x_i, \lambda)|.$$

Формальное применение метода наименьших квадратов, как уже отмечалось выше, недопустимо. Всегда надо помнить, что это – аппарат линейного

регрессионного анализа, требующий строгости применения регрессионных моделей. Формальные модели (без оглядки на физические свойства оригинала) не всегда применимы даже в области полученных экспериментальных данных (см. примеры 2 и 3 ниже).

Точки **поляры самолета** – зависимости $c_{xa} = f(c_{ya})$ получают из комплекса расчетов, экспериментов в аэродинамических трубах и в натуральных полетах. Они имеют достаточно большой разброс и требуют обработки для того, чтобы можно было построить какие-либо детерминированные модели. Т.е. необходима аппроксимация. Прежде всего встает вопрос, с помощью какого класса (вида) функций следует проводить эту аппроксимацию: \ln , \sin , \exp , линейные, квадратичные, полиномы или др. Исходя из физических соображений, первые 4 вида функций следует признать неприемлемыми. При сравнении квадратичной и полиномиальной аппроксимации на первый взгляд кажется, что чем выше степень полинома, тем лучше – тем ближе к экспериментальным точкам будет проходить кривая. Однако, оказывается, что это далеко не всегда так! Полиномы высоких степеней могут иметь точки перегиба, которых по физическому смыслу поляры быть не должно. Полином выше второй степени все-таки можно использовать для аппроксимации поляры, но только такой, который в интересующей нас области обладает следующим набором свойств:

- не имеет корней,
- не имеет точек перегиба,
- вблизи $c_{ya} = 0$ имеет точку минимума.

Еще более глубокие требования обнаруживаются при аппроксимации зависимости c_{xa} от числа M . В итоге применения метода наименьших квадратов и выбора наилучшего из приемлемых полиномов поляра самолета Ту-154Б в пределах 3 % погрешности аппроксимируется двумерным полиномом, содержащим M в степенях 0 и 4, а c_{ya} в степенях 0, 1, 2, 3. Аппроксимация поляры самолета Ту-134А содержащим M в степенях 0, 4, 8, а c_{ya} в степенях 0, 1, 2, 3.

В одной из разработок гражданской авиации приводилась функциональная зависимость стоимости пассажирского самолета от его вместимости. Никаких сведений о допустимом диапазоне применения этой формулы и использованном статистическом материале не давалось. При попытке использовать указанную зависимость в комплексе с другими экономическими, прочностными, аэродинамическими расчетами для параметрических исследований перспективного типа самолета на 700 пассажиров получались бессмысленные результаты. При пристальном изучении этой зависимости было обнаружено, что она имеет вид параболы выпуклостью вверх. Само по себе это нужное свойство, так как с ростом пассажироместимости стоимость самолета растет все медленнее. Однако у такой параболы есть максимум, который должен находиться **вне** зоны возможных расчетов. У предложенной зависимости максимум оказался в районе 300 пассажиров, что привело к **отрицательной стоимости** самолета на 700 пассажиров. Ситуация похожа на приведенную на рис. 54. Очевидно, что такой зависимостью можно было пользоваться только в области собранного статистического материала, т.е. в то время в диапазоне масс самолетов от Ан-2 до Ил-62.

В конце 80-х годов XX века проводились обширные исследования коррозионных повреждений силовых элементов и обшивки самолетов. Собранный статистический материал по своим формальным признакам позволял произвести регрессионный анализ зависимости коррозионных повреждений от календарного срока службы самолета, что и было сделано в одном из учреждений гражданской авиации. На рис. 55 представлен пример такого материала для листов обшивки самолетов Ан-24, причем в указанной работе была построена **формальная** линейная регрессия, призванная отображать **рост** процента листов

обшивки, поврежденных коррозией, от календарного срока службы самолета t (длинный пунктир): $32,1 - 0,2269 \cdot (t - 138,6)$. Очевидно, что такой "зависимостью" пользоваться нельзя ни для наблюдавшегося диапазона налетов, ни для прогнозов, так как она убывающая.

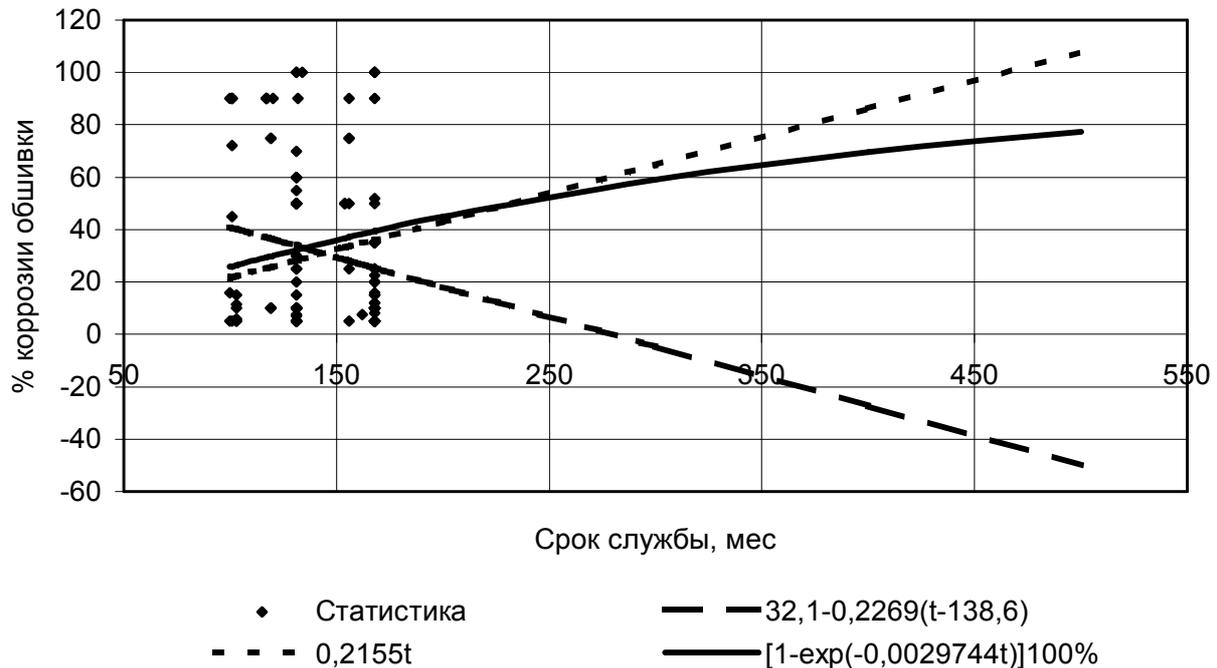


Рис. 55.

Попытка построить линейную же регрессию, имеющую один естественный признак: прохождение **через начало координат** (новый самолет не имеет коррозии) – привела к регрессии без свободного члена: $0,2155 \cdot t$. Условие **возрастания** тоже было выполнено. Но эта зависимость не может дать вызывающего доверие прогноза: при каком сроке службы будет достигнут порог в 80 % листов обшивки, поврежденных коррозией. Ведь с таким же успехом по ней можно спрогнозировать и 120 % поврежденных листов!

Только серьезный **физичный** подход к выбору аппроксимирующей функции может дать осмысленный прогноз. Указанная физическая зависимость должна иметь следующие свойства:

- проходить через начало координат;
- быть возрастающей;
- быть выпуклой вверх;
- асимптотически приближаться к 100 % при $t \rightarrow \infty$.

Таким свойствам удовлетворяет найденная регрессионная модель вида: $100\% \cdot (1 - e^{-0,0029744t})$, позволяющая сделать вывод о том, что средний процент листов обшивки, поврежденных коррозией, превысит 80 %-ый порог после 40 лет службы самолета.

6.5. Понятие о конфлюэнтном анализе

Основой конфлюэнтного анализа является представление **структуры** исследуемых случайных величин в виде двух составляющих (*конфлюэнтная модель*):

$$\xi = a + \xi', \quad \eta = b + \eta',$$

где a и b математические ожидания – структурные компоненты, а ξ' и η' – стохастические компоненты (случайные) с **нулевым** математическим ожиданием. Предполагается, что связь между факторами ξ и η определяется связью структурных компонент, а стохастические имеют характер шума. В конфлюэнтном анализе определяются оценки для параметров λ в соотношении связи: $\varphi(a, b, \lambda) = 0$ – структурном уравнении, а также законы распределения стохастических компонент. Его удобно применять, когда исследуется связь зашумленных параметров, например, погрешностями измерения, определяемыми ненаблюдаемыми факторами.

Нетрудно видеть, что постановка задачи рассмотренного в § 6.4 регрессионного анализа является **частным случаем** конфлюэнтного, когда дисперсия "аргумента" равна нулю: $D(\xi) = D(\xi') = 0$. С другой стороны, отыскание распределения стохастических компонент представляет собой типичную задачу корреляционного анализа. Таким образом, конфлюэнтный анализ можно рассматривать, как обобщение корреляционного и регрессионного.

В предыдущем параграфе было отмечено, что регрессионный анализ не претерпевает принципиальных изменений, если допустить случайность "аргумента". Т.е. структурирование факторов не мешает установлению регрессии между ними. Более того, регрессия как раз и представляет собой структурное уравнение, связывающее математические ожидания факторов. Из этого становится ясно, почему понятие конфлюэнтного анализа не прижилось на практике: вместо него обычно говорят о совместном проведении корреляционного и регрессионного анализа или даже только о корреляционном анализе, подразумевая регрессионный его частью. В любом случае каких-либо специфических методов конфлюэнтного анализа нет – он основывается на ранее изученных в §§ 6.2 и 6.4 методах.

6.6. Непрерывные случайные величины и понятие о теории фильтрации

До сих пор мы рассматривали дискретные случайные величины и возможные методы обработки информации о наблюдении за ними. Ясно, что эти же методы применимы и к непрерывным случайным величинам, если вычислять начальные и центральные моменты по формулам:

$$\nu_r(\xi) = \int_{-\infty}^{\infty} \xi^r f(\xi) d\xi, \quad \mu_r(\xi) = \int_{-\infty}^{\infty} (\xi - a)^r f(\xi) d\xi,$$

а под операторами математического ожидания и дисперсии понимать:

$$a = E(\xi) = \int_{-\infty}^{\infty} \xi f(\xi) d\xi, \quad \sigma^2 = D(\xi) = \int_{-\infty}^{\infty} (\xi - a)^2 f(\xi) d\xi,$$

где $f(\xi)$ – закон распределения (плотность распределения вероятностей).

Однако у случайных **процессов**, описываемых непрерывными величинами, зависящими от времени, есть существенная особенность: их значения в любые два момента времени не только коррелированы, но и физически зависи-

мы. Так, например, если снимать данные с записей параметров полета через достаточно большие промежутки времени, то можно считать их некоррелированными. (**Что** именно можно считать достаточно большим промежутком времени, необходимо решать в каждом конкретном случае отдельно.) Обычно запись делается с малым интервалом времени, позволяющим отследить **развитие** процесса, но тогда результаты необходимо считать коррелированными.

Степень их коррелированности можно оценить *ковариацией* (§ 5.1), однако более употребительно понятие *автокорреляционной функции*: $R(\tau) \equiv E\{x(t - \tau) \cdot x(t)\}$, основанное не на центральном, а на начальном моменте. Эта функция характеризует ковариацию (коррелированность, связанность) **значений** параметра, **отстоящих** на интервал времени τ . Очевидное неравенство нулю автокорреляционной функции не позволяет применять дисперсионный анализ непосредственно к результатам измерений. В этом случае приходится представлять их через специфические независимые функции.

Рассмотрим **стационарный** случайный процесс по некоторому параметру t . Данные наблюдений за ним всегда можно представить в виде конечной суммы ряда Фурье:

$$x(t) = \sum_{j=1}^k (a_j \cos \omega_j t + b_j \sin \omega_j t),$$

где a_j и b_j – независимые, **центрированные** (с математическим ожиданием равным нулю), нормально распределенные величины с дисперсиями

$$\sigma^2 \{a_j\} = \sigma^2 \{b_j\} = \sigma_j^2,$$

не зависящими от времени. Вместе с независимыми тригонометрическими функциями они и составляют множество **независимых** величин, заменяющих наблюдаемые значения $x(t)$. Тогда нетрудно показать, что автокорреляционная функция зависит только от интервала времени τ и обладает свойством четности:

$$R(\tau) = R(-\tau) = \sum_{j=1}^k \sigma_j^2 \cos \omega_j \tau.$$

Таким образом, наблюдение за случайным процессом можно **заменить** наблюдением за дисперсиями $\sigma_j^2(\omega)$ – амплитудами колебаний с частотами ω_j . Эти величины составляют так называемый **спектр** процесса – понятие особенно близкое физикам, привыкшим мыслить в терминах гармонических колебаний. (Рассмотренная процедура "замены" является одним из следствий известной в теории преобразований Фурье теоремы Хинчина – Винера.)

Другими задачами статистического анализа непрерывных факторов являются задачи, связанные с исследованием **зашумленных** факторов. В этих задачах основное внимание сосредоточивается не на отдельных замерах, а на правильном представлении процесса в развитии. А при такой постановке вопроса корреляция из нежелательной становится полезной! С ее помощью можно попытаться **отделить** полезный сигнал от помех. Так, например, в записях

установившегося полета сразу можно отсеять те отдельные точки, в которых скорость самолета существенно отличается. Этот тривиальный пример и является, по сути, фильтрацией информации.

В 30-х годах XX века А.Н. Колмогоров и Н. Винер независимо друг от друга поставили перед собой именно такую задачу: разработать математический аппарат для обоснованного выравнивания случайных процессов. Их разработки стали впоследствии называться теорией фильтрации.

При решении таких практических задач, как наведение летательных аппаратов, обеспечение мягкой посадки, распределение ресурсов, идентификация переменных параметров систем, обнаруживается **недостаток информации**. Он связан не столько с влиянием неизвестных факторов, сколько с пробелами и возмущениями во временном ряду наблюдаемых факторов. Поэтому основной задачей теории фильтрации считается **выделение** исходного сигнала из искаженной или неполной информации.

Пусть $x(t)$ – неискаженный исходный сигнал, а $\eta(t)$ – случайная аддитивная помеха, т.е. наблюдать можно только величину $x(t) + \eta(t)$. Родоначальники теории фильтрации сформулировали задачу отыскания такой функции $\psi(t)$, называемой *фильтром*, которая позволила бы с учетом **предыстории** восстановить информацию о сигнале с помощью преобразования:

$$y(t) = \int_{-\infty}^t \psi(t - \tau)[x(\tau) + \eta(\tau)]d\tau.$$

Конечно, такое восстановление не может быть полным и выход фильтра не совпадает с исходным сигналом, т.е. $y(t) \neq x(t)$. Поэтому должен быть принят какой-то рецепт для того, **что** именно можно считать восстановлением информации. А.Н. Колмогоров и Н. Винер предложили в качестве **критерия близости** $y(t)$ к $x(t)$ использовать минимум квадрата рассогласования между ними.

Следует, однако, заметить, что это рассогласование невозможно вычислить, так как $x(t)$ неизвестно. Кроме того, реально построенный фильтр не сможет реализовывать функцию вида $\psi(t)$. На самом деле эта функция характеризуется вектором собственных параметров γ и может быть реализована с неизбежными помехами в виде случайных факторов ω . Т.е. фильтр представляет собой функцию вида $\psi(t, \omega, \gamma)$, а выход фильтра – функцию $y(t, \omega, \gamma)$.

Для того, чтобы решить такую сложную оптимизационную задачу и определить неискаженный исходный сигнал, необходимо не только подобрать наилучший фильтр (т.е. его оптимальные параметры γ), но и **исключить** влияние случайных факторов ω .

Параметры γ фильтра должны быть оптимальными не в каждый отдельный момент времени t , а на всем интересующем нас интервале времени T , поэтому имеет смысл определять квадрат рассогласования в виде осредненной по этому интервалу величины:

$$\sigma^2(\omega, \gamma) = \frac{1}{T} \int_0^T [y(t, \omega, \gamma) - x(t)]^2 dt.$$

Исключить влияние ω в этом критерии оптимальности можно, если осреднить эту величину по вероятности на всем возможном интервале случайных факторов:

$$I(\gamma) = \int_{-\infty}^{\infty} \sigma^2(\omega, \gamma) \cdot p(\omega) \cdot d\omega \Rightarrow \min.$$

Именно эта величина и принимается в качестве **критерия** оптимальности в задаче отыскания **наилучшего фильтра**. Плотность распределения вероятностей $p(\omega)$ подбирается в процессе предварительного изучения системы.

Р. Калман и Р. Бьюси решили частную задачу в предположении:

$$y = H \cdot x + \eta \quad \text{при} \quad \dot{x} = Ax + B\xi, \quad x(t = t_0) = x_0,$$

где ξ – случайное воздействие. Решение этой задачи выражается не аналитически, а дискретным алгоритмом – *фильтром Калмана*, в процессе последовательного пошагового уточнения дающим **несмещенную** оценку параметра системы.

Раздел 3 ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Глава 7. Статистические методы планирования эксперимента

7.1. Проблемы построения эксперимента

Авиационная и космическая техника относится к "плохо организованным системам". Большое количество факторов, существенно влияющих на результат работы этой техники, требует построения специфических моделей, как, например, аэродинамические характеристики летательного аппарата, которые представляют собой не что иное, как регрессионные модели. Их применение диктуется способностью описать сложные логические связи, без четкого и аргументированного понимания которых невозможно правильно эксплуатировать технику. Их разработка требует постановки и проведения сложных и дорогостоящих экспериментов (например, летных испытаний и продувок в аэродинамических трубах) с исследованием влияния множества параметров. Для этого необходимо планирование эксперимента – совокупность действий, объединенных **целью** исследования и направленных на разработку **стратегии** экспериментирования от начальных до заключительных этапов изучения объекта исследований (от получения априорной информации до создания работоспособной математической модели или определения оптимальных условий) [12].

С другой стороны, между объектом исследования и результатами эксперимента, полученными и обработанными в целях этого исследования, существует огромная дистанция. Дело не только в том, что эксперимент может быть поставлен разным образом, а его результаты обработаны различными методами. Дело еще в том, что объект исследования – *оригинал* – многообразен и бесконечно сложен в своих проявлениях, только часть из которых мы можем наблюдать в эксперименте – *модели*. Т.е. необходимо четко себе представлять, что результаты эксперимента – это лишь более или менее представительная **выборка** при наблюдении за генеральной совокупностью. А это требует использования методов математической статистики для проведения успешного анализа результатов эксперимента. Особую роль здесь играют *дисперсионные модели* (*модели дисперсионного анализа*, § 6.3) и *регрессионные модели* (*модели регрессионного анализа*, § 6.4).

Ситуация еще более усложняется, если исследуется не единичный объект, а сложная система. В этом случае наибольшим затруднением становится множественность целей подсистем и элементов, а также множественность критериев оценки их работы (*многокритериальность*). Различные критерии могут вступать в противоречие друг с другом, однако даже их частичное игнорирование может полностью исказить систему как целое.

Эффективность авиаремонтного завода (АРЗ) не может быть оценена каким-либо единым показателем, например, доходом. АРЗ работает не в изолированном пространстве, а в среде авиакомпаний, которые противятся росту цены ремонта. По-видимому, кроме дохода необходимо учитывать и себестоимость, и производительность труда, и общий объем работ, и что-то еще. С другой стороны, для цеха ремонта гидросистем придется учитывать и другие дополнительные факторы. Так, например, резкое увеличение производительности труда приведет к простоям из-за отсутствия фронта работ. Таким образом, при неправильной организации труда (игнорировании множественности задач и критериев деятельности подсистем) безусловно прогрессивный рост производительности труда может обернуться общим проигрышем.

Изучение "*плохо организованных систем*" требует построения сложных моделей, которые в свою очередь, требуют обширных экспериментальных исследований. Возникает необходимость разработки такой постановки эксперимента, которая обеспечила бы выполнение сразу нескольких условий: практическую осуществимость (быстрота, экономичность), достаточную полноту (для обоснования выводов), наглядность (удобство интерпретации).

В преодолении указанных трудностей помогают методы планирования эксперимента. С их помощью отнюдь не производится жесткое планирование всего до мелочей – это невозможно и неразумно хотя бы потому, что в жизни возможны отклонения. Здесь широко применяется последовательный анализ, возврат к предыдущему эксперименту для его уточнения или переориентации и другие приемы, приводящие в конечном итоге к построению необходимой модели. Единого рецепта здесь нет и быть не может – каждую исследуемую систему с учетом ее конкретики можно разделить на более простые подсистемы и рассматривать их во взаимодействии. Однако есть математически обоснованные методы планирования эксперимента для построения статистических математических моделей. Они объединяются в классическую теорию эксперимента.

В теории *планирования эксперимента* сформировалась система основных терминов и понятий. Прежде всего имеется в виду, что изучаемая система может быть представлена "черным ящиком" (рис. 51 § 6.1). Кроме того, вводятся следующие определения.

Эксперимент – это система операций, воздействий и (или) наблюдений, направленных на получение информации об объекте при исследовательских испытаниях [12].

Опыт – это отдельная часть эксперимента, воспроизводящая исследуемое явление в определенных задаваемых условиях при возможности регистрации его результатов.

Пассивный эксперимент – эксперимент в отсутствии управляемых факторов: выходные факторы зависят только от неуправляемых входных и неконтролируемых факторов (шума).

Активный эксперимент – эксперимент в отсутствии неуправляемых входных факторов: выходные факторы зависят только от управляемых входных и неконтролируемых факторов (шума).

Теория планирования эксперимента занимается именно активным экспериментом и предполагает планирование значений управляемых входных фак-

торов \mathbf{X} в области их возможных значений. Таким образом, планом эксперимента является некоторая **совокупность уровней факторов \mathbf{X}** , построенная для определенных целей исследования. Обычно он изображается в виде матрицы плана, в которой строки отвечают опытам, а столбцы – факторам, т.е. элемент X_{ij} матрицы плана обозначает определенный уровень j -го фактора в i -м опыте:

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & X_{Nk} \end{pmatrix}.$$

С точки зрения унификации изображения планов экспериментов такая форма наиболее предпочтительна. Основными ее чертами являются:

- дробление на элементарные опыты (испытания) и расположение их по строкам;
- указание для каждого опыта (в каждой строке) реализуемых в нем уровней всех факторов (в строке перечислены уровни всех исследуемых факторов, реализуемые в данном опыте);
- расположение факторов по столбцам (в столбце перечислены уровни одного фактора, встречающиеся во всех опытах).

По количеству исследуемых факторов и их уровней эксперименты имеют свою устоявшуюся терминологию. Так, k -факторным называется эксперимент, в котором результат рассматривается в зависимости от изменения k контролируемых управляемых входных факторов (вектор \mathbf{X} имеет размерность k). Если все исследуемые факторы имеют в эксперименте одинаковое количество h уровней, то такой эксперимент называется h -уровневым.

Для наиболее экономичного получения достоверного результата, удовлетворяющего заданным требованиям, необходимо решить ряд проблем постановки эксперимента:

- 1) собственно эксперимент:
 - а) формулировка целей и задач эксперимента,
 - б) выбор наблюдаемого выходного фактора,
 - в) выбор управляемых факторов,
 - г) выбор уровней этих факторов (количественных или качественных, фиксированных или случайных),
 - д) подбор сочетаний уровней факторов;
- 2) планирование эксперимента:
 - а) определение необходимого числа опытов,
 - б) определение порядка проведения отдельных опытов,
 - в) выбор метода рандомизации,
 - г) составление математической модели для описания результатов;
- 3) анализ результатов эксперимента:
 - а) сбор и обработка данных,

- б) вычисление статистик (выборочных функций) для проверки гипотез,
- в) интерпретация результатов эксперимента.

Для успешного решения первой из этих проблем применяются методы **теории моделирования**, третьей – методы **обработки информации**. Методам решения второй проблемы и частично первой посвящена теория **планирования эксперимента**.

Основные принципы планирования эксперимента позволяют сделать эксперимент практически реализуемым и информативным.

1. Принцип отказа от полного перебора возможных входных состояний. Для получения исчерпывающей информации о поведении выходных факторов Y , вообще говоря, необходимо проведение такого числа опытов в эксперименте, чтобы все уровни каждого из входных факторов встречались друг с другом во всех возможных комбинациях. Понятно, что с ростом числа факторов и уровней объем такого эксперимента быстро увеличивается, а для непрерывно изменяющегося хотя бы одного фактора становится бесконечным. Поэтому от полного перебора отказываются сознательно, строго отбирая лишь существенные факторы и вводя дискретную сетку уровней непрерывных факторов, исходя из необходимого вида математической модели.

2. Принцип последовательного усложнения математической модели (принцип последовательного планирования). Планирование эксперимента не следует понимать, как безусловно однократную завершённую операцию. При постановке сложного эксперимента бывает весьма полезно пользоваться приемами *последовательного анализа*, следуя Е.С. Вентцель [8]: "Основной принцип теории планирования эксперимента состоит в том, что любое принятое заранее решение должно пересматриваться с учетом полученной новой информации".

Исследования основных, наиболее значительных свойств объекта начинают с двухуровневого эксперимента, которому соответствует линейная регрессионная модель. В этом случае и эксперимент минимален по объему, и основные характеристики объекта (возрастание, убывание) можно выявить. После этого строится более сложная (например, квадратичная, требующая не менее трех уровней) модель, исходя из задачи исследований, по тем факторам, по которым это необходимо. Проведение опытов для новых уровней факторов дополняет информацию, при этом результаты предыдущих опытов не пропадают, а используются полностью для построения новой модели. В этой процедуре возможно и даже необходимо пользоваться проверкой "качества" полученных на каждом этапе *моделей* с помощью проверки их *адекватности* и усложнять их именно там, где это требуется.

3. Принцип сопоставления с шумом. Точность регрессионной модели должна сопоставляться с интенсивностью случайного неконтролируемого воздействия. Действительно: бессмысленно ставить дорогостоящий эксперимент для получения точной модели, если результаты эксперимента сами по себе обладают большой итоговой погрешностью. Конечно, повторные опыты могут существенно увеличить точность эксперимента, но здесь имеется в виду именно итоговая погрешность всего эксперимента с учетом повторения опытов.

4. Принцип рандомизации (принцип приведения к случайности влияния факторов). Рандомизация – это обеспечение случайности влияния действующих на систему факторов, не поддающихся или поддающихся с трудом учету и контролю. В рандомизированной системе влияние действующих факторов можно считать вполне случайным, что позволяет учитывать их статистически. Простейший прием рандомизации плана эксперимента состоит в случайном переборе уровней факторов. Понятие рандомизации ввел Фишер, заметивший, что разделение общей дисперсии на частные, соответствующие исследуемым факторам, тем эффективнее, чем случайнее вклад каждого фактора. Этот же вывод можно отнести и к неучтенным на первых порах факторам, что подталкивает к учету как можно большего числа факторов. Т.е. наблюдается полная противоположность "хорошо организованным системам", где эксперимент надо было "очистить" от "посторонних", "мешающих" факторов. Иными словами, *рандомизация* – это искусственный перевод неслучайных факторов, которые **невозможно учесть** в эксперименте, в разряд случайных, что позволяет избавиться от систематических погрешностей, связанных с этими факторами.

5. Принцип оптимальности плана (наличие критерия оптимальности). Цели исследований можно достичь только в таком эксперименте, который практически осуществим, достаточно информативен и нагляден. Все это требует применения оптимизации, т.е. формулировки *критерия оптимальности* (§ 4.5). При выработке критерия оптимальности плана эксперимента безусловно следует учесть и минимизацию объема и стоимости эксперимента, и максимизацию объема и эффективности (в статистическом смысле) получаемой информации. К последним в современном планировании эксперимента принято относить критерии оптимальности планов для регрессионного анализа. Вообще говоря, даже для одной и той же задачи могут быть сформулированы несколько критериев оптимальности. Из нескольких десятков таких критериев отметим наиболее распространенные, связанные с точностью оценивания коэффициентов регрессии:

- критерий D-оптимальности обеспечивает минимизацию обобщенной дисперсии оценок коэффициентов регрессионной модели (минимизацию объема эллипсоида рассеяния);
- критерий A-оптимальности обеспечивает минимизацию суммы дисперсий оценок коэффициентов регрессионной модели;
- критерий E-оптимальности обеспечивает минимизацию наибольшей оси эллипсоида рассеяния.

Математический аппарат поиска оптимальных планов достаточно сложен. Аналитическое решение такой задачи возможно только в простейших частных случаях. Поэтому на практике оно реализуется численными методами на ЭВМ.

7.2. Назначение плана эксперимента

В § 2.2 целью планирования эксперимента было сформулировано получение максимума достоверной информации при минимуме затрат. Под затрата-

ми здесь могут пониматься не только материальные затраты непосредственно на эксперимент (чаще всего этот вопрос сводится к планированию объема эксперимента – см. § 7.3), но и затраты времени и ресурсов.

При планировании эксперимента следует иметь в виду, что информация должна быть получена *достоверной* и сформулированной наиболее компактно. Эти условия можно выполнить только с помощью дисперсионного и регрессионного анализа, т.е. с помощью подбора *дисперсионных* и *регрессионных* математических *моделей* (см. §§ 6.3, 6.4), наилучшим образом соответствующих результатам эксперимента. Т.е. необходимо получение таких оценок параметров моделей, статистически определенных по результатам эксперимента, которые обладают свойствами *эффективности, состоятельности и несмещенности* (§ 5.3).

Эффективность статистической оценки, как известно, тем выше, чем меньше ее дисперсия. А поскольку дисперсия оценки любого параметра зависит от множества факторов: условий постановки эксперимента, его погрешности и, как видно из нижеприведенных примеров, от его плана, то очевидна возможность управления общей дисперсией через дисперсии от влияния этих отдельных факторов.

Перечисленные требования к эксперименту, конкретизированные в терминах математической статистики, позволяют осознанно выбирать наиболее удачный план с точки зрения применения тех или иных методов статистического анализа.

В том случае, когда весы имеют погрешность, наиболее естественный и традиционный способ взвешивания трех предметов А, В, С заключается в предварительном "холостом взвешивании" пустой чашки весов и последующем взвешивании отдельных предметов. План такого эксперимента показан в табл. 14. Здесь матрицей плана является матрица из +1 и -1, представляющих два уровня каждого фактора: $x_{ij} = +1$ и $x_{ij} = -1$ (для каждого предмета это означает присутствие или отсутствие его на чашке весов).

В итоге вес А определится как $y_1 - y_0$, а дисперсия этого результата формулой:

$$\sigma^2\{A\} = \sigma^2\{y_1 - y_0\} = 2\sigma^2\{y\},$$

где $\sigma\{y\}$ – погрешность однократного взвешивания.

Этот же эксперимент можно провести по плану, изображенному в табл. 15. В нем "холостое взвешивание" заменено взвешиванием всех трех предметов вместе. В этом случае вес А определится выражением: $\frac{1}{2}(y_1 - y_2 - y_3 + y_4)$, а дисперсия

$$\sigma^2\{A\} = \sigma^2\left\{\frac{1}{2}(y_1 - y_2 - y_3 + y_4)\right\} = \frac{1}{4}4\sigma^2\{y\} = \sigma^2\{y\}$$

окажется вдвое меньше, т.е. результат **вдвое точнее!** Такой эффект достигнут за счет того, что каждый вес теперь получается в результате четырех взвешиваний, а не двух. Иными словами, для того, чтобы определить вес каждого предмета по первому плану требуется произ-

Таблица 14.

| № опыта | А | В | С | результат |
|---------|----|----|----|-----------|
| 0 | -1 | -1 | -1 | y_0 |
| 1 | +1 | -1 | -1 | y_1 |
| 2 | -1 | +1 | -1 | y_2 |
| 3 | -1 | -1 | +1 | y_3 |

Таблица 15.

| № опыта | А | В | С | результат |
|---------|----|----|----|-----------|
| 0 | +1 | +1 | +1 | y_4 |
| 1 | +1 | -1 | -1 | y_1 |
| 2 | -1 | +1 | -1 | y_2 |
| 3 | -1 | -1 | +1 | y_3 |

вести только два опыта: замер с пустой чашей весов и с эти предметом, зато по второму плану все четыре опыта участвуют в определении веса каждого предмета. Таким образом, стоящее в знаменателе формулы дисперсии число степеней свободы выросло с 2 до 4. В этом смысле первый план фактически распадается на три отдельных плана однофакторных двухуровневых экспериментов, а второй оказывается планом трехфакторного двухуровневого эксперимента.

При изучении систем с большим количеством факторов приходится иметь дело с понятием полноты плана, отражающим перебор всех уровней исследуемых факторов. Поэтому введем это понятие. Полным факторным экспериментом называется такой эксперимент, план которого обеспечивает полный перебор всевозможных сочетаний всех уровней и всех факторов. План полного факторного эксперимента называют полным планом. Объем полного плана (т.е. количество необходимых опытов) легко подсчитать:

$$h_1 \times h_2 \times \dots \times h_k = \prod_{j=1}^k h_j,$$

где k – число факторов, а h_j – число уровней каждого j -го фактора. Так, например, планы табл. 14 и 15 неполные.

Если *уровни* исследуемых факторов распределены в планах **симметрично**, т.е. все встречаются одинаковое число раз, то такие планы называются сбалансированными. Симметрия эта не геометрическая, а "смысловая": если в одном опыте встречается один уровень данного фактора, то в каком-то другом опыте должен встретиться и другой. Например, если есть опыт ABC, то должны быть опыты и ВСА, и САВ, что обеспечит симметрию опытов по использованию уровней А, В, С факторов. Так, например, часть плана табл. 15, включающая строки 1, 2, 3, можно считать сбалансированной, так как в каждом опыте присутствует только один фактор, принимающий уровень +1, в этом смысле все факторы здесь присутствуют симметрично.

Нетрудно доказать, что *полный план* всегда *сбалансирован*.

Очевидна сложность применения полных планов в случае большого количества факторов и уровней. В этом случае строят неполные планы. Однако произвольный неполный план может не выполнять принцип *рандомизации*, поэтому его приходится строить с некоторыми **признаками** полноты и симметрии. Чаще всего для обеспечения рандомизированности неполных планов они разбиваются на группы опытов. Такие группы опытов в плане, объединенных каким-либо общим свойством, называются блоками. Например, сбалансированность внутри блока приводит к сбалансированным блокам, полнота – к полным блокам. Если полноту блока нельзя отметить, как его свойство, то план будет называться *неполноблочным*. Так, например, часть плана табл. 15, включающую строки 1, 2, 3, можно считать сбалансированным неполным блоком.

Нередко возникает задача распределения k уровней k факторов в k опытах. Эту задачу позволяют решить латинские квадраты. В латинских квадратах размером $k \times k$ в каждой строке и в каждом столбце размещаются только **разнотипные** элементы и только по **одному** разу. В табл. 16 приведены примеры

возможных латинских квадратов размерами 3×3 и 5×5 . Все латинские квадраты можно построить из одного простой перестановкой строк и столбцов. Самый простой способ построения исходного для такого процесса квадрата состоит в выборе случайного порядка элементов первой строки, а следующие строки получают последовательным сдвигом элементов, чем обеспечивается неповторимость и строк, и столбцов.

Таблица 16.

| | | |
|---|---|---|
| A | B | C |
| B | C | A |
| C | A | B |

| | | | | |
|---|---|---|---|---|
| A | B | C | D | E |
| B | C | D | E | A |
| C | D | E | A | B |
| D | E | A | B | C |
| E | A | B | C | D |

7.3. Планирование объема эксперимента

В данном параграфе будут рассмотрены подходы к определению частного показателя планов эксперимента – его объема.

Для ясности изложения (но, не умаляя общности) допустим, что эксперимент имеет целью определение с контролируемой погрешностью значения единственного параметра x наблюдаемого явления. Изложим по порядку усложнения различные методы определения объема плана такого эксперимента.

A. Простейший подход к планированию объема эксперимента выражается известной поговоркой: "Семь раз отмерь, один – отрежь!". Математическое обоснование этой житейской мудрости ясно из следующего.

Поскольку погрешность связана с рассеиванием параметра, т.е. с отклонением этой случайной величины от своего математического ожидания, то простейший подход к оценке погрешности можно свести к точечной оценке рассеивания. На языке математической статистики это выражается известным соотношением:

$$\sigma\{\bar{x}\} = \frac{\sigma}{\sqrt{N}},$$

где σ – среднее квадратическое отклонение контролируемого параметра x в **одиночном** опыте (измерении), \bar{x} – его среднее выборочное значение по N одинаковым опытам, $\sigma\{\bar{x}\}$ – среднее квадратическое отклонение среднего выборочного.

Итак, если априори известно (например, в виде паспортных данных измерительной техники) значение σ контролируемого параметра при одиночном опыте, то вышеприведенная формула позволяет определить необходимое число опытов для обеспечения требуемой погрешности δ :

$$N = \frac{\sigma^2}{\delta^2},$$

при этом **не существует закон** распределения x .

Б. Точечные оценки контролируемого параметра обладают всеми необходимыми свойствами несмещенности, состоятельности, эффективности и достаточности только при больших объемах эксперимента. Кроме того, определение объема эксперимента по дисперсии не дает вероятностной оценки обеспечения заданной погрешности. А такая оценка необходима, поскольку результаты любого эксперимента следует рассматривать как выборку из генеральной совокупности проявления исследуемого процесса. Поэтому более качественную оценку требуемого объема эксперимента следует проводить с помощью доверительных интервалов (§ 5.5). Однако для их построения необходимо знать закон распределения этого параметра; в отличие от предыдущих методов – это и есть "плата" за качество оценки.

Поскольку контролируемый параметр чаще всего оценивают с помощью выборочного среднего, постольку для оценки его точности можно использовать доверительный интервал для математического ожидания этого контролируемого параметра. Этот интервал обычно строится симметрично относительно выборочного среднего, и его радиус непосредственно определяет точность оценки (вернее, *погрешность*) при заданной "уверенности" – доверительной вероятности γ . Так, например, для **нормально** распределенного параметра погрешность его средней выборочной оценки (радиус доверительного интервала, половина его) определяется одной из формул:

$$\delta = u_{0,5\gamma} \frac{\sigma}{\sqrt{N}} \quad \text{или} \quad \delta = t_{\gamma} \frac{s}{\sqrt{N}},$$

(первой – при априори известном значении σ по 4-й строке табл. 10 § 5.4, второй – при неизвестном σ , но найденной выборочной несмещенной его оценке – s по 5-й строке табл. 10 § 5.4), где $u_{0,5\gamma}$ – аргумент стандартизованной функции нормального закона распределения [18], определяемый из условия: $\Phi(u_{0,5\gamma}) = 0,5 \cdot \gamma$ (см. § 5.5), где γ – назначенная доверительная вероятность; а t_{γ} – из таблицы распределения Стьюдента с $N - 1$ степенями свободы при вероятности $p = 1 - \gamma$. Отсюда вытекают формулы для определения объема эксперимента, необходимого для обеспечения погрешности δ с доверительной вероятностью γ :

$$N = \frac{u_{0,5\gamma}^2 \sigma^2}{\delta^2} \quad \text{или} \quad N = \frac{t_{\gamma}^2 s^2}{\delta^2}.$$

В общем случае контролируемой величины, распределенной по другому закону (не нормальному), следует исходить из общего определения доверительного интервала и строить его на основании именно этого закона распределения. Можно также пользоваться приближенными законами распределения.

В. Близок по смыслу к изложенному "простейший критериальный" подход к определению объема эксперимента. Пусть x – **нормально** распределенный параметр, о значении которого судят по найденному выборочному среднему \bar{x} из эксперимента объемом N . Критериальный подход (§ 5.6) позволяет с помощью ре-

результатов этого эксперимента проверить гипотезу $H_0: a = a_0$ о том, что математическое ожидание a имеет значение a_0 при разнообразных конкурирующих гипотезах. Это бывает необходимо, когда оценка сугубо несимметрична.

Из 4-ой и 5-ой строк табл. 10 § 5.4 известно, что величины:

$$u = \frac{\bar{x} - a_0}{\sigma / \sqrt{N}} \quad \text{и} \quad t = \frac{\bar{x} - a_0}{s / \sqrt{N}}$$

имеют нормальное распределение и, соответственно, распределение Стьюдента с $N - 1$ степенями свободы. Эти величины выступают в качестве критерия при проверке гипотезы $H_0: a = a_0$: если критерий больше критического значения, то проверяемая гипотеза отвергается, если меньше – принимается (говорят, что по результатам выборки нет оснований отвергнуть проверяемую гипотезу). В табл. 17 приводится сводка окончательных формул для определения объема эксперимента, необходимого для проверки соответствующих конкурирующих гипотез, полученных с помощью аппарата, изложенного в § 5.6.

Таблица 17.

| Конкурирующая гипотеза | Требуемый объем эксперимента | |
|--|--|---|
| | при известном σ | при неизвестном σ |
| $a \neq a_0$ | $u_{\frac{1-\alpha}{2}}^2 \times \frac{\sigma^2}{(\bar{x} - a_0)^2}$ | $t_{1-\frac{\alpha}{2}}^2 \times \frac{s^2}{(\bar{x} - a_0)^2}$ |
| $a > a_0$ | $u_{\frac{1-\alpha}{2}}^2 \times \frac{\sigma^2}{(\bar{x} - a_0)^2}$ | $t_{1-\alpha}^2 \times \frac{s^2}{(\bar{x} - a_0)^2}$ |
| $a < a_0$ | $u_{\frac{1-\alpha}{2}}^2 \times \frac{\sigma^2}{(\bar{x} - a_0)^2}$ | $t_{\alpha}^2 \times \frac{s^2}{(\bar{x} - a_0)^2}$ |
| $a = a_1$ | $(u_{\frac{1-\beta}{2}} + u_{\frac{1-\alpha}{2}})^2 \times \frac{\sigma^2}{(a_1 - a_0)^2}$ | $(t_{1-\beta} + t_{1-\alpha})^2 \times \frac{s^2}{(a_1 - a_0)^2}$ |
| <p>Проверяемая гипотеза $a = a_0$
 α – вероятность ошибки I рода (отвергнуть верную гипотезу)
 β – вероятность ошибки II рода (принять неверную гипотезу)
 Значения вероятностей γ в функции u_γ соответствуют таблице стандартизованного нормального закона распределения вероятностей из [18].</p> | | |

Однако на практике целесообразнее пользоваться не этими конечными формулами, а логикой построения доверительных интервалов, так как постановка реальных задач бывает весьма далека от постановки задачи проверки статистических критериев.

Требуется с помощью летных испытаний выяснить возможность приема самолета нового типа на аэродроме с располагаемой посадочной дистанцией $L_{a/d} = 1500$ м. Разброс посадочных дистанций на самолете-прототипе составляет 300 м. Сколько посадок необходимо произвести в летных испытаниях, чтобы обеспечить ответ на поставленный вопрос с погрешностью $\delta = 50$ м при допустимой вероятности ошибки $\alpha = 0,1$ %?

Прежде всего, необходимо иметь оценку среднего квадратического отклонения σ при одном опыте (посадке). Проводить летные испытания для получения этой оценки слишком дорого и небезопасно, поэтому воспользуемся сведениями о прототипе, перенося результат на исследуемый тип самолета. Однако этот обычный прием следует делать аккуратно. Предположим, что посадочная дистанция распределена по нормальному закону. На самом деле это не так, потому что отрицательных значений посадочной дистанции или очень больших не бывает. Но в первом приближении это можно допустить, так как предполагаемый центр распределения с учетом разброса (от 1350 м до 1650 м) существенно удален от начала координат. В таком случае воспользуемся правилом 3σ (вероятность попадания нормально распределенной случайной величины в интервал радиусом 3σ вокруг математического ожидания равна 0,9973, что легко проверить по таблицам функции Лапласа), что даст оценку среднего квадратического отклонения при одной посадке $\sigma \approx 50$ м, исходя из разброса для прототипа, равного 300 м, т.е. двух радиусов по 3σ .

1 подход – "тривиальный" согласно рассмотренному выше методу А. Трактовка σ , непосредственно как погрешности, очевидно неприемлема, так как из вышеприведенных рассуждений следует явная недостаточность $\sigma \approx 50$ м для практического использования. Для уменьшения среднего квадратического отклонения средней посадочной дистанции по результатам N посадок до значения, соответствующего уменьшению разброса от 300 м до 100 м (погрешность 50 м), т.е. в 3 раза необходимо произвести: $N = 3^2 = 9$ посадок. Именно тогда $3\sigma(N)$ составят величину не более допустимой погрешности 50 м.

Исходя из правила 3σ , вероятность ошибки при этом можно оценить величиной 0,27 %, что существенно хуже требуемой вероятности ошибки.

2 подход – "формальный доверительный" согласно рассмотренному выше методу Б. Применение симметричного доверительного интервала для обеспечения заданной погрешности (δ – радиус интервала) при заданной доверительной вероятности приводит к соотношению:

$$N > \frac{u_{\gamma}^2 \sigma^2}{\delta^2} = \frac{3,29^2 50^2}{50^2} = 10,82.$$

Напомним, что в этом случае $\gamma = 0,5$ ($1 - \alpha$) для таблицы [18]. Таким образом, при таком подходе требуется провести летный эксперимент из 11 посадок.

3 подход – "неформальный доверительный". Обратим внимание на то, что отклонения значений посадочной дистанции в меньшую сторону для целей нашей практической задачи несущественны. Это означает, что достаточно построить односторонний доверительный интервал: от $-\infty$ до $L_{a/d}$. Не входящая в него часть числовой оси правее $L_{a/d}$ (рис. 56) представляет собой область риска принять неверное решение. Таким образом, вероятность попадания истинного значения посадочной дистанции в эту область не должна превышать $\alpha = 0,1$ %, т.е. для таблицы [18] следует принять $\gamma = 0,5 - \alpha$. Предполагая, что центр распределения расположен левее $L_{a/d}$ на 50 м (для обеспечения запаса), т.е. $a_0 = L_{\text{пос}} = L_{a/d} - \delta$, среднее арифметическое значение \bar{L} посадочных дистанций при N посадках должно удовлетворять условию:

$$P\{\bar{L} > a_0 + \delta = L_{a/d}\} = 0,001, \quad \text{т.е.} \quad N > \frac{3,09^2 50^2}{50^2} = 9,55.$$

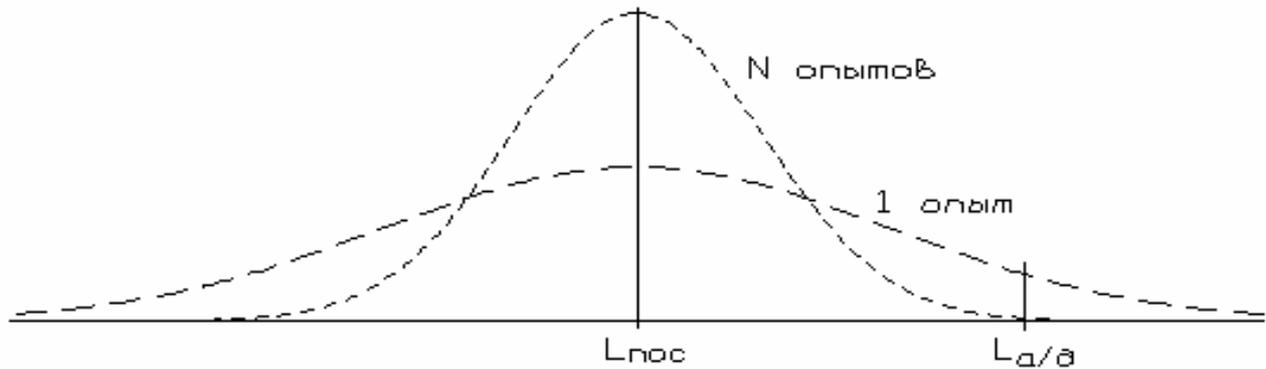


Рис. 56.

Таким образом, при этом подходе необходимо произвести 10 посадок.

4 подход – "формальный критериальный" согласно рассмотренному выше методу **В**.

В этом случае проверяемая гипотеза: $a_0 = L_{\text{пос}} = L_{a/d} - \delta$ (для обеспечения запаса). При конкурирующей гипотезе о том, что истинное значение посадочной дистанции отличается от допустимой a_0 в любую сторону более, чем на требуемую погрешность: $|L_{\text{пос}} - a_0| > \delta$, руководствуемся первой строкой табл. 17. Тогда необходимое число посадок составит (при $\alpha = 0,001$): $N > 10,82$, как и при "формальном доверительном" 2-м подходе.

5 подход – "неформальный критериальный". В этом случае та же проверяемая гипотеза: $a_0 = L_{\text{пос}} = L_{a/d} - \delta$. При конкурирующей гипотезе о том, что истинное значение посадочной дистанции превосходит допустимую a_0 более, чем на требуемую погрешность: $L_{\text{пос}} - a_0 > \delta$, руководствуемся второй строкой табл. 17. Тогда необходимое число посадок составит (при $\alpha = 0,001$): $N > 9,55$, как и при "неформальном доверительном" 3-м подходе.

6 подход. Попробуем получить ответ на поставленный вопрос с помощью **одного** единственного испытания, в результате которого получено значение $L_{\text{пос}} = x_1 = 1410$ м. В этом случае естественно выдвинуть гипотезу $H_0: a = a_0 = x_1 = 1410$ м. С выдвижением альтернативной гипотезы дело обстоит сложнее. На первый взгляд в качестве альтернативной гипотезы следует принять $H_1: a > x_1 = 1410$ м. Тогда при заданной вероятности ошибки $\alpha = 0,001$ (которую резонно считать ошибкой I рода) напрашивается использование близости распределения этой величины к нормальному с известным средним квадратическим отклонением $\sigma \approx 50$ м. По таблице функции Лапласа [18] для стандартизованной величины $\frac{x^* - a_0}{\sigma}$ находим ее значение 3,09, соответствующее вероятности $0,5 - \alpha = 0,499$. Итак, гра-

ница критической области определится соотношением: $x^* = 3,09 \cdot \sigma + x_1 = 1565$ м. В этом случае расположения критической области (рис. 46) при очевидном $1410 \text{ м} < 1565 \text{ м}$ следует вывод о приемлемости гипотезы H_0 , т.е. о возможности эксплуатации этого типа самолета на данном аэродроме.

Однакостораживают промежуточные результаты: граница критической области находится в зоне недопустимых значений $x^* = 1565 \text{ м} > L_{a/d} = 1500 \text{ м}$! Так как требования к точности и достоверности, т.е. δ и α , заданы, то это кажущееся несоответствие может быть ликвидировано только одним способом: заменой статистического материала, на основании которого и делался вывод. Однако выясним, нельзя ли разрешить эту проблему только лишь заменой альтернативной гипотезы.

7 подход. В рассматриваемом примере логически возможно выдвижение только одной гипотезы другого вида: $H_1: a + a_1$. Остаются под вопросом лишь значения a_1 и β .

Проанализируем требование точности. Заданная погрешность $\delta = 50$ м может интерпретироваться, как величина уверенного (с некоторой вероятностью) различения двух значе-

ний $L_{\text{пос}}$. Тогда в качестве a_1 следовало бы рассматривать для единственного имеющегося результата $L_{\text{пос}} = x_1 = 1410$ м значение: $a_1 = x_1 + \delta = (1410 + 50) \text{ м} = 1460 \text{ м}$.

Выбор допустимой вероятности ошибки второго рода β (ошибочного принятия неверной гипотезы) следует делать из соображений более опасных ее последствий, т.е. следует назначить β существенно меньше, чем α , например, $\beta = 0,0001$. Это и понятно: ущерб от принятия данного аэродрома к эксплуатации при его непригодности к этому неизмеримо больше. При этом погрешность $\delta = 50$ м по-прежнему рассматривается как расстояние между центрами двух однотипных распределений (a_0 и a_1 на рис. 47), а граница критической области x^* должна находиться между ними:

$$u_{0,5-\alpha} = \frac{x^* - a_0}{\sigma} \quad \text{и} \quad u_{0,5-\beta} = \frac{a_1 - x^*}{\sigma}.$$

Однако, нетрудно видеть, что эти соотношения невыполнимы: по таблице функции Лапласа $u_{0,5-\alpha} = 3,09$, $u_{0,5-\beta} = 3,72$, а правые части принимают значения, безусловно меньшие 1, так как x^* располагается на отрезке длиной 50 м, чему равен знаменатель. Этот факт свидетельствует о невозможности сделать положительный вывод по имеющимся данным и при самой сложной альтернативной гипотезе. Необходимо собрать другой статистический материал.

8 подход. Рассмотрим в качестве такого материала данные N посадок, а в качестве оценки для $L_{\text{пос}}$ используем среднее по этим результатам \bar{L} . Тогда последние соотношения из предыдущего подхода примут вид:

$$u_{0,5-\alpha} = \frac{x^* - a_0}{\sigma / \sqrt{N}} \quad \text{и} \quad u_{0,5-\beta} = \frac{a_1 - x^*}{\sigma / \sqrt{N}},$$

и из них, исключая x^* , можно получить выражения для необходимого объема эксперимента N :

$$N > (u_{0,5-\beta} + u_{0,5-\alpha})^2 \times \frac{\sigma^2}{(a_1 - a_0)^2},$$

а исключая σ – положение границы критической области, но только после задания a_0 :

$$x^* = \frac{u_{0,5-\alpha} \cdot a_1 + u_{0,5-\beta} \cdot a_0}{u_{0,5-\alpha} + u_{0,5-\beta}}.$$

Расчеты для рассматриваемого примера дают $N > 46,38$. На практике следовало бы по результатам 47 посадок вычислить среднюю величину посадочной дистанции \bar{L} и принять ее в качестве a_0 , выдвинув тем самым гипотезу $H_0: a = a_0$. Если вычисленная после этого граница критической области x^* окажется правее полученной величины \bar{L} и левее $L_{a/d} - \delta$, то не будет оснований отвергать гипотезу H_0 , т.е. можно разрешить эксплуатацию нового типа самолета на данном аэродроме.

На практике при выборе приема планирования объема эксперимента следует исходить каждый раз только из условий конкретной задачи исследований. Так, например, при отсутствии возможности достаточно уверенного определения или оценки σ применение подходов с 3-го по 8-ой видится неоправданно сложным. Не стоит также забывать об экономических аспектах и приемах последовательного анализа (§ 5.6).

7.4. Планирование однофакторного эксперимента

В классической теории эксперимента рассматриваются планы, составленные с целью проведения дисперсионного анализа определенного вида. Поэтому естественно, что исходной посылкой служит цель исследования. Исходя из нее, формируется представление о виде предполагаемой математической модели явления, числе исследуемых факторов и их уровнях (см. § 6.1).

Простейшим в этом контексте следует считать *однофакторный эксперимент*, который призван установить факт зависимости результата эксперимента от значения исследуемого **единственного** фактора, т.е. влияние других факторов предполагается несущественным. Рассмотрим подробно последовательность действий для построения плана такого эксперимента на конкретном примере.

Исследование влияния срока хранения на качество ГСМ.

Подразумевается существование такого единственного измеримого параметра y , который характеризует качество ГСМ и который может зависеть от некоторых входных факторов X . Так как измерительная аппаратура имеет погрешность, то для исследования влияния срока хранения на параметр y необходимо производить многократные рандомизированные его замеры при различных сроках хранения, например, по n замеров в течение нескольких (m) месяцев. (Задавшись определенными требованиями к достоверности выводов, с помощью методик, рассмотренных в § 7.3, можно определить необходимый объем эксперимента, т.е. значения n и m .) Весь массив данных в таком случае удобно сформировать в виде (табл. 18) матрицы результатов эксперимента $Y = \{y_{ji}\}$, где $j = 1, 2, \dots, m$ – номер значения (уровня) исследуемого фактора (срока хранения – месяца), а $i = 1, 2, \dots, n$ – порядковый номер замера качества y (в течение каждого месяца). Таким образом, планируемый эксперимент оказывается многоуровневым – m -уровневым.

Таблица 18.

| № замера (i) | Значения фактора – уровни (j) | | | |
|--------------------|-----------------------------------|-------------|-----|-------------|
| | 1 | 2 | ... | m |
| 1 | y_{11} | y_{21} | ... | y_{m1} |
| 2 | y_{12} | y_{22} | ... | y_{m2} |
| ... | ... | ... | ... | ... |
| n | y_{1n} | y_{2n} | ... | y_{mn} |
| Средние по уровням | \bar{y}_1 | \bar{y}_2 | ... | \bar{y}_m |

Для последующего дисперсионного анализа необходимо задать вид математической модели. Он определяется исследуемым явлением и той гипотезой, которая проверяется. Выдвинем гипотезу о том, что срок хранения T влияет на характеристику качества ГСМ y . Тогда естественно представить себе математическую модель для дисперсионного анализа (*дисперсионную модель*) исследуемого влияния в виде:

$$y_{ji} = \mu + T_j + \varepsilon_{ji},$$

где μ – математическое ожидание генерального значения параметра y за весь исследуемый срок, T_j – добавок к μ от влияния исследуемого фактора (срока хранения в течении j месяцев), ε_{ji} – нормально распределенная погрешность в каждом единичном опыте (i -ом замере в j -ом месяце).

В **унифицированном** виде, принятом в планировании эксперимента, матрица такого **плана** (табл. 19) содержит $n \times m$ строк, соответствующих единичным опытам.

| № опыта | Уровни фактора | Результат замера |
|--------------|-----------------------|------------------|
| 1 | 1 | y_{11} |
| 2 | 1 | y_{12} |
| ... | ... | ... |
| n | 1 | y_{1n} |
| $n + 1$ | 2 | y_{21} |
| ... | ... | ... |
| $m \times n$ | m | y_{mn} |

Этот план полностью *рандомизированного однофакторного многоуровневого эксперимента* позволяет произвести дисперсионный анализ (§ 6.3) и выявить влияние исследуемого единственного фактора. Для этого, как это принято в дисперсионном анализе, разбивают статистическую оценку общей дисперсии:

$$s^2 = \frac{1}{mn - 1} \sum_{j=1}^m \sum_{i=1}^n (y_{ji} - \bar{\bar{y}})^2,$$

где

$$\bar{\bar{y}} = \frac{1}{m} \sum_{j=1}^m \bar{y}_j, \quad \text{а} \quad \bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ji},$$

на две части: **остаточную** – внутреннюю, отражающую разброс результатов единичных опытов вокруг средних по группам экспериментальных данных каждого месяца, обусловленную неконтролируемой погрешностью эксперимента:

$$s_0^2 = \frac{1}{m(n - 1)} \cdot \sum_{j=1}^m \sum_{i=1}^n (y_{ji} - \bar{y}_j)^2,$$

и **межгрупповую**, отражающую разброс средних (месячных) экспериментальных данных между собой из-за **влияния исследуемого фактора**:

$$s_A^2 = \frac{n}{m - 1} \sum_{j=1}^m (\bar{y}_j - \bar{\bar{y}})^2.$$

Далее оценку значимости фактора Т (зависимости качества ГСМ от срока хранения) получают с помощью критерия Фишера для сравнения двух дисперсий при уровне значимости α (строка 13 табл. 10). Если

$$\frac{s_A^2}{s_0^2} > F_{1-\alpha} [m - 1, m(n - 1)],$$

то значимое превышение s_A^2 над s_0^2 свидетельствует о значимости фактора Т, т.е. о существовании выраженной зависимости у от Т (качества ГСМ от срока хранения). В этом случае можно оценить дисперсию влияния исследуемого фактора величиной:

$$\frac{s_A^2 - s_0^2}{n}.$$

Если наоборот:

$$\frac{s_0^2}{s_A^2} > F_{1-\alpha} [m(n - 1), m - 1],$$

то s_A^2 значимо меньше s_0^2 , т.е. случайной погрешности эксперимента, поэтому влияние фактора Т на у несущественно и им можно пренебречь. В этом случае в качестве оценки погрешности (дисперсии) следует использовать выражение общей дисперсии s^2 , имеющей

число степеней свободы больше, чем S_A^2 и S_0^2 . В случае, когда нельзя говорить о значимом превосходстве одной из дисперсий над другой, влияние исследуемого входного фактора сравнимо с погрешностью эксперимента или влиянием неучтенных факторов, поэтому конкретный вывод невозможен.

7.5. Планирование двухфакторного эксперимента

План двухфакторного эксперимента рассмотрим на примере задачи об испытаниях шин.

Пусть для испытания выделено 12 шин новой марки и 4 типа автомобиля: I, II, III, IV; а технические условия таковы, что перестановка колес во время испытания исключается.

В этом эксперименте предполагается наличие двух существенных факторов: T – автомобиль и S – место установки шины на автомобиле, распознаваемое по условным номерам: 1, 2, 3, 4. Таким образом, количество уровней каждого фактора оказывается равным 4. Легко видеть, что рассматриваемые факторы независимы, поэтому можно не исследовать третий фактор их совместного влияния. Математическая *дисперсионная модель* такого эксперимента может быть записана в виде:

$$y_{ijk} = \mu + T_j + S_i + \varepsilon_{ijk},$$

где y_{ijk} – результат испытания шины на i -м месте j -го автомобиля; μ – средний ожидаемый результат (математическое ожидание); T_j – отклонение от μ , обусловленное влиянием типа автомобиля; S_i – отклонение от $\mu + T_j$, обусловленное влиянием места установки шины на автомобиле; ε_{ijk} – нормально распределенная погрешность оценки результата эксперимента, обусловленная влиянием неучтенных факторов.

Естественное стремление поставить эксперимент наиболее полным образом для данного примера означает, что надо было бы испытать шины на каждом месте всех автомобилей. Т.е. провести полный факторный эксперимент, обеспечивающий всевозможные сочетания всех уровней и всех факторов, объемом

$$\prod_{i=1}^k h_i = 4 \times 4 = 16$$

опытов, где $k = 2$ – число факторов, а $h_i = 4$ – число уровней каждого фактора. Наличие только 12 шин не позволяет построить полный план такого эксперимента. Как же следует разместить шины с наибольшей выгодой для наибольшей представительности результатов?

Если попытаться разместить шины, полностью укомплектовав первые 3 типа автомобилей, то последний окажется не у дел. В целях экономии средств это может быть и выгодно, но не позволит получить информацию об эксплуатационных свойствах шин на четвертом типе автомобиля. Однако с формальной точки зрения такой способ обладает существенным преимуществом перед другими – полнотой. Мы получаем *полный план двухфакторного эксперимента* для 3 уровней фактора типа автомобиля и 4 уровней фактора места расположения шин. В этом примере очевидно, что полнота плана без дополнительных усилий влечет за собой его сбалансированность – симметричность распределения отдельных опытов в эксперименте, т.е. в нашем примере расположение шин на каждом из трех автомобилей симметрично. Кроме того, нет нужды заботиться о рандомизации – в рамках рассматриваемой задачи различие между шинами не исследуется, а сбалансированность полного плана ставит исследуемые факторы в равные условия.

Аналогичные рассуждения можно провести для размещения шин на трех определенных местах (например, за исключением 4-го) всех 4 автомобилей.

Эти неудачные попытки наводят на мысль о неизбежности построения *неполного* плана. Симметричное размещение 12 шин на 4 местах 4 автомобилей возможно только группами – *блоками* (сериями). Блоки можно рассматривать как по автомобилям, так и по местам расположения. В нашем примере блоки не могут быть *полными* – шины будут испытываться не на каждом месте каждого автомобиля: план эксперимента оказывается *неполноблочным*.

Стремление получить как можно более полную информацию подталкивает к построению *сбалансированных блоков*: на каждом автомобиле разместить по 3 шины так, чтобы любые два места расположения встречались одинаковое число раз. А рандомизацию исследуемых факторов можно обеспечить случайным распределением незанятых номеров мест между автомобилями или полным перебором таковых.

Таким образом, мы приходим к *неполноблочному сбалансированному плану двухфакторного четырехуровневого рандомизированного эксперимента*, изображенному в табл. 20. В ней номер строки (места) можно интерпретировать, как номер отдельного опыта для определения результата испытаний шин на данном месте установки. Аналогичные рассуждения допустимы и относительно номера столбца (автомобиля). Т.е. возникает множество возможностей изображать один и тот же план одного и того же эксперимента.

Исходным в нашем примере является представление об опыте, связанном с испытанием одной шины. Поэтому более естественным представляется изображение плана эксперимента в **унифицированной** форме, показанной в табл. 21. Матрица этого плана выглядит несколько непривычно для своего "названия", однако, она дает четкое изображение уровней всех исследуемых факторов для каждого опыта (т.е. где устанавливается каждая шина).

Таблица 20.

| № места (i) | № автомобиля (j) | | | | |
|-------------|------------------|----------------|----------------|----------------|----------------|
| | I | II | III | IV | |
| 1 | +1 | -1 | +1 | +1 | S ₁ |
| 2 | -1 | +1 | +1 | +1 | S ₂ |
| 3 | +1 | +1 | +1 | -1 | S ₃ |
| 4 | +1 | +1 | -1 | +1 | S ₄ |
| | T ₁ | T ₂ | T ₃ | T ₄ | μ |

Таблица 21.

| № опыта (шины) | Факторы | |
|----------------|--------------|---------|
| | № автомобиля | № места |
| 1 | I | 1 |
| 2 | I | 3 |
| 3 | I | 4 |
| 4 | II | 2 |
| 5 | II | 3 |
| 6 | II | 4 |
| 7 | III | 1 |
| 8 | III | 2 |
| 9 | III | 3 |
| 10 | IV | 1 |
| 11 | IV | 2 |
| 12 | IV | 4 |

7.6. Планирование многофакторного эксперимента

Рассмотрим расширение примера предыдущего параграфа на случай трех факторов.

Для испытаний выделено по 4 шины 4 марок А, В, С, D и 4 типа автомобиля: I, II, III, IV. Перестановка колес во время испытаний исключается.

В этом эксперименте присутствуют 3 существенных фактора: Т – тип автомобиля, S – место установки шины на автомобиле, распознаваемое по условным номерам: 1, 2, 3, 4, и Q – тип шины. Как и в предыдущем параграфе, количество уровней каждого фактора равно 4 и

факторы независимы. Математическая *дисперсионная модель* такого эксперимента, не исследующего совместное влияние факторов, может быть записана в виде:

$$Y_{ijmk} = \mu + T_j + S_i + Q_m + \varepsilon_{ijmk},$$

где Y_{ijmk} – результат испытания шины m -ой марки на i -м месте j -го автомобиля; Q_m – отклонение от $\mu + T_j + S_i$ (см. § 7.5), обусловленное влиянием m -й марки шины; ε_{ijmk} – погрешность эксперимента, обусловленная влиянием неучтенных факторов.

Полный факторный эксперимент в этом случае требует $4 \times 4 \times 4 = 64$ опыта, т.е. 64 шины. В нашем распоряжении их только 16, что позволяет полностью обеспечить ими все автомобили.

Поэтому, несмотря на то, что весь план неполный, блоки в нем можно сделать полными и рандомизированными, т.е. установить на каждый автомобиль в случайном порядке все четыре типа шин. В таком случае составить сбалансированный рандомизированный план помогают *латинские квадраты* (см. § 7.2). Такой план в двух разных видах представлен в табл. 22 и 23.

Таблица 22.

| № места (i) | № автомобиля (j) | | | |
|-------------|------------------|----|-----|----|
| | I | II | III | IV |
| 1 | A | B | C | D |
| 2 | D | A | B | C |
| 3 | C | D | A | B |
| 4 | B | C | D | A |

Таблица 23.

| № опыта (шины) | Факторы | | |
|----------------|--------------|---------|--------------|
| | № автомобиля | № места | № марки шины |
| 1 | I | 1 | A |
| 2 | I | 2 | D |
| 3 | I | 3 | C |
| 4 | I | 4 | B |
| 5 | II | 1 | B |
| 6 | II | 2 | A |
| 7 | II | 3 | D |
| 8 | II | 4 | C |
| 9 | III | 1 | C |
| 10 | III | 2 | B |
| 11 | III | 3 | A |
| 12 | III | 4 | D |
| 13 | IV | 1 | D |
| 14 | IV | 2 | C |
| 15 | IV | 3 | B |
| 16 | IV | 4 | A |

Как видно из рассмотренных примеров, образное представление многофакторных планов с числом уровней больше 2, весьма затруднено. Поэтому для **изучения** особенностей таких планов будем в дальнейшем рассматривать только **двухуровневые факторы**.

Для наглядности редуцируем рассмотренный пример трехфакторного четырехуровневого эксперимента к двум уровням по каждому фактору. Т.е. рассмотрим 2 типа шин А и В, 2 типа автомобилей I и II с 2 местами для испытания: 1 и 2. План в этом случае выглядит так, как показано в табл. 24. Этот же план в унифицированном виде представлен в табл. 25, где x_1 обозначает фактор Т (тип автомобиля), x_2 – фактор S (место установки шины на автомобиле), а x_3 – Q (тип шины), +1 первые уровни всех факторов (1, I, А) и –1 – вторые уровни (2, II, В).

Таблица 24.

| № места (i) | № автомобиля (j) | |
|-------------|------------------|----|
| | I | II |
| 1 | А | В |
| 2 | В | А |

Таблица 25.

| № опыта (i) | Факторы (j) | | | Результат опыта (i) |
|-------------|-------------|-------|-------|---------------------|
| | x_1 | x_2 | x_3 | |
| 1 | +1 | +1 | +1 | y_1 |
| 2 | +1 | -1 | -1 | y_2 |
| 3 | -1 | +1 | -1 | y_3 |
| 4 | -1 | -1 | +1 | y_4 |

Не случайно план табл. 25 совпадает с планом табл. 15 § 7.2 в задаче о взвешивании. И тот и другой являются *сбалансированным полноблочным планом трехфакторного двухуровневого, полностью рандомизированного эксперимента для дисперсионного анализа*.

Нетрудно видеть, что план эксперимента табл. 25, обеспечивающий получение результатов y_1, y_2, y_3 и y_4 , предоставляет принципиальную возможность провести **и другие** виды статистического анализа: регрессионный и корреляционный. Прежде, чем перейти к обсуждению этой возможности, введем некоторые определения, характеризующие **свойства** плана.

Обозначим элементы матрицы плана через x_{ij} . Тогда для столбцов $j = 1, 2, 3$ легко убедиться в следующем:

1) сумма элементов любого столбца равна нулю:

$$\sum_{i=1}^N x_{ij} = 0,$$

где N – число опытов; это свойство говорит о том, что план симметричный;

2) сумма квадратов элементов столбца (длина столбца) на единицу больше числа k факторов:

$$\sum_{i=1}^N x_{ij}^2 = k + 1,$$

– это условие нормировки, а $k + 1$ называется числом степеней свободы плана;

3) скалярное произведение любых двух различных столбцов плана равно нулю:

$$\sum_{i=1}^N x_{im} x_{in} = 0 \quad \text{при } m \neq n,$$

это свойство носит название ортогональности (оно приводит к **независимости** определения всех коэффициентов линейной регрессионной модели);

4) число опытов N совпадает с числом коэффициентов уравнения линейной регрессии $k + 1$ – в этом случае план называется насыщенным для выбранной **линейной регрессионной** модели. Действительно: по 4 результатам эксперимента можно определить все 4 коэффициента линейной по трем факторам регрессии исследуемого параметра y :

$$y = \lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3.$$

Следует обратить внимание на то, что понятие насыщенности **связано** именно с линейной регрессионной математической моделью исследуемого явления.

В случае $N > k + 1$ план ненасыщенный – определение всех коэффициентов линейной регрессионной модели не исчерпывает собранную информацию, не использует все степени свободы плана. В этом случае кроме коэффициентов регрессии можно получить **дополнительные** статистические оценки, например, проверить гипотезу о равенстве нулю коэффициента регрессии совместного влияния пары факторов, или построить доверительные интервалы для коэффициентов регрессии, или провести параллельно дисперсионный анализ. При насыщенном плане это невозможно.

В случае $N < k + 1$ план сверхнасыщенный и позволяет получить оценки только **части** коэффициентов линейной регрессии.

Заметим, что оценки для коэффициентов линейной регрессии $\tilde{\lambda}_j = \frac{1}{N} \sum_{i=1}^N x_{ij} y_i$ получаются с дисперсией $\sigma^2 \{\tilde{\lambda}_j\} = \frac{1}{N} \sigma^2 \{y\}$, в N раз меньшей, чем

результат каждого опыта. Для определения $\tilde{\lambda}_0$ можно воспользоваться теми же соображениями, что и для других коэффициентов. Однако получающаяся формула свободного члена $\tilde{\lambda}_0 = \frac{1}{N} \sum_{i=1}^N y_i$ нарушает единообразие. Поэтому модернизиру-

ем план за счет дополнительного фиктивного фактора x_0 "прочих" влияний, которые всегда присутствуют, при всех опытах эксперимента. Новый вид плана представлен в табл. 26, где фигурной скобкой помечен собственно план. Симметричная матрица, представляющая этот план, в которой первый столбец, первая строка и главная диагональ состоят из +1, а все остальные элементы –1, называется матрицей Адамара.

Таблица 26.

| № опыта (i) | Факторы (j) | | | | Результат опыта (i) |
|-------------|-------------|-------|-------|-------|---------------------|
| | x_0 | x_1 | x_2 | x_3 | |
| 1 | +1 | +1 | +1 | +1 | y_1 |
| 2 | +1 | +1 | -1 | -1 | y_2 |
| 3 | +1 | -1 | +1 | -1 | y_3 |
| 4 | +1 | -1 | -1 | +1 | y_4 |
| | | план | | | |

В заключение данного параграфа отметим, что составленный в табл. 26 *сбалансированный полноблочный план трехфакторного двухуровневого, полностью рандомизированного эксперимента для дисперсионного анализа* обладает свойствами симметричности, ортогональности и насыщенности для **линейной регрессионной модели**, для него также выполнено условие нормировки. Однако этот план не является полным, так как вместо

$$\prod_{j=1}^k h_j = 2^3 = 8$$

опытов содержит только 4. Но, несмотря на это, план позволяет провести **линейный** регрессионный анализ по всем **трем** факторам.

7.7. Неполные и неортогональные планы

В конце предыдущего параграфа было отмечено, что план эксперимента по табл. 26 неполный, хотя и позволяет построить **линейную** регрессионную модель, являясь для нее *насыщенным*.

Дополним его, для чего добавим всевозможные строки, в которых уровень +1 встречается дважды, и строку со всеми -1. В табл. 27 показан такой план.

Таблица 27.

| № опыта (i) | Факторы (j) | | | | Результат опыта |
|-------------|----------------|---|----------------|----------------|-----------------|
| | x ₀ | x ₁ | x ₂ | x ₃ | |
| 1 | +1 | +1 | +1 | +1 | У ₁ |
| 2 | +1 | +1 | -1 | -1 | У ₂ |
| 3 | +1 | -1 | +1 | -1 | У ₃ |
| 4 | +1 | -1 | -1 | +1 | У ₄ |
| 5 | +1 | +1 | +1 | -1 | У ₅ |
| 6 | +1 | +1 | -1 | +1 | У ₆ |
| 7 | +1 | -1 | +1 | +1 | У ₇ |
| 8 | +1 | -1 | -1 | -1 | У ₈ |
| | | 
план | | | |

Вообще говоря, полный план можно составлять, руководствуясь следующим **правилом**, обеспечивающим полный перебор всевозможных комбинаций двух уровней основных факторов (кроме фиктивного x₀). В первом столбце знаки меняют через один. Во втором знаки встречаются парами, т.е. чередуются через 2. В третьем – четверками, чередуясь через 4. Далее, если необходимо – через следующие степени двойки. Построенный

по этому правилу полный план всегда обладает свойствами *симметричности* и *ортогональности*, что можно проверить непосредственно. Нетрудно видеть, что план табл. 27 соответствует этому правилу построения, если в нем изменить порядок строк (опытов) на следующий: 1, 7, 6, 4, 5, 3, 2, 8.

Новый план стал полным – его объем стал равняться $2^3 = 8$, чего мы и добивались, но теперь нарушилось свойство насыщенности для линейной регрессии. Для нее *план* стал *ненасыщенным*. Следовательно, есть возможность расширения регрессионной модели на столько членов, на сколько объем эксперимента (число опытов) N стал больше числа степеней свободы k + 1. В нашем случае N = 8, а k + 1 = 4, т.е. можно добавить 4 таких новых члена модели, которые не изменят числа **основных** факторов. Таким образом, в модель можно добавить всевозможные **сочетания** (произведения) основных факторов и рассматривать ее в следующем, уже нелинейном, виде:

$$y = \lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \lambda_{12} x_1 x_2 + \lambda_{13} x_1 x_3 + \lambda_{23} x_2 x_3 + \lambda_{123} x_1 x_2 x_3.$$

Под такую **расширенную** регрессионную модель можно построить *насыщенный план*. В табл. 28 представлен такой план, полученный из полного

плана для линейной трехфакторной регрессии (перекомпонованный план табл. 27) **добавлением** столбцов сочетаний (произведений) основных факторов.

Таблица 28.

| № опыта (i) | Факторы (j) | | | | | | | | Результат опыта |
|-------------|----------------|----------------|----------------|----------------|-------------------------------|-------------------------------|-------------------------------|--|-----------------|
| | x ₀ | x ₁ | x ₂ | x ₃ | x ₁ x ₂ | x ₁ x ₃ | x ₂ x ₃ | x ₁ x ₂ x ₃ | |
| 1 | +1 | +1 | +1 | +1 | +1 | +1 | +1 | +1 | Y ₁ |
| 2 | +1 | -1 | +1 | +1 | -1 | -1 | +1 | -1 | Y ₂ |
| 3 | +1 | +1 | -1 | +1 | -1 | +1 | -1 | -1 | Y ₃ |
| 4 | +1 | -1 | -1 | +1 | +1 | -1 | -1 | +1 | Y ₄ |
| 5 | +1 | +1 | +1 | -1 | +1 | -1 | -1 | -1 | Y ₅ |
| 6 | +1 | -1 | +1 | -1 | -1 | +1 | -1 | +1 | Y ₆ |
| 7 | +1 | +1 | -1 | -1 | -1 | -1 | +1 | +1 | Y ₇ |
| 8 | +1 | -1 | -1 | -1 | +1 | +1 | +1 | -1 | Y ₈ |
| | | план | | | | | | | |

Легко проверить, что построенный таким образом *насыщенный* план обладает всеми рассмотренными в § 7.6 свойствами. Условие **нормировки** выполняется для всех столбцов очевидным образом. **Симметричность** столбцов парных произведений следует из условия ортогональности основных столбцов. В симметричности столбца тройного произведения можно убедиться с помощью следующих рассуждений, основанных на виде основных столбцов, позволяющем последовательно разбивать суммы на две части:

$$\sum_{i=1}^N x_{i1}x_{i2}x_{i3} = \sum_{i=1}^{N/2} x_{i1}x_{i2}x_{i3} + \sum_{i=N/2+1}^N x_{i1}x_{i2}x_{i3} = \sum_{i=1}^{N/2} x_{i1}x_{i2} \cdot 1 - \sum_{i=N/2+1}^N x_{i1}x_{i2} \cdot 1 = 0,$$

$$\text{так как} \quad \sum_{i=1}^{N/2} x_{i1}x_{i2} = \sum_{i=N/2+1}^N x_{i1}x_{i2}.$$

Ортогональность столбцов расширения и основных столбцов следует при $k \neq l \neq m \neq k$ из симметричности столбца тройного произведения:

$$\sum_{i=1}^N (x_k)_i (x_l x_m)_i = \sum_{i=1}^N x_{ik} x_{il} x_{im} = 0,$$

при $l = k$ – из симметричности и нормированности основных столбцов:

$$\sum_{i=1}^N (x_k)_i (x_k x_m)_i = \sum_{i=1}^N x_{ik} x_{ik} x_{im} = \sum_{i=1}^N 1 \cdot x_{im} = 0,$$

а для тройного произведения:

$$\sum_{i=1}^N (x_k)_i (x_k x_l x_m)_i = \sum_{i=1}^N x_{ik} x_{ik} x_{il} x_{im} = \sum_{i=1}^N 1 \cdot x_{il} x_{im} = 0$$

– из ортогональности основных столбцов. Ортогональность столбцов расширения между собой проверяется аналогично:

$$\sum_{i=1}^N (x_k x_l)_i (x_l x_m)_i = \sum_{i=1}^N x_{ik} x_{il} x_{il} x_{im} = \sum_{i=1}^N 1 \cdot x_{ik} x_{im} = 0,$$

$$\sum_{i=1}^N (x_k x_l)_i (x_k x_l x_m)_i = \sum_{i=1}^N x_{ik} x_{il} x_{ik} x_{il} x_{im} = \sum_{i=1}^N 1 \cdot 1 \cdot x_{im} = 0.$$

Все это вместе взятое, позволяет рассматривать описанный способ построения насыщенного плана 3-факторного эксперимента для **расширенной нелинейной** регрессионной модели, как способ построения насыщенного плана для любой 7-факторной **линейной** регрессионной модели вида:

$$y = \lambda_0 + \sum_{j=1}^7 \lambda_j x_j.$$

Действительно: его насыщенность позволяет определить все 8 коэффициентов такой линейной модели, несмотря на неполноту. Полный план 7-факторного эксперимента должен был бы содержать $\prod_{j=1}^k h_j = 2^7 = 128$ опытов!

Нельзя не прийти к выводу о неприемлемости реализации полных планов для многофакторных экспериментов.

Проведенный переход от плана табл. 26 к плану табл. 28 показывает возможность построения неполного, но **насыщенного** плана для эксперимента с большим числом факторов, исходя из плана с меньшим числом факторов. Так, при переходе от неполного плана табл. 26 к полному плану табл. 27 было замечено, что исходный план составляет симметричную, ортогональную часть от полного плана, насыщенную для **линейной регрессионной модели**. Переход от полного плана трехфакторного эксперимента табл. 27 к неполному плану семифакторного эксперимента табл. 28 позволил построить тоже неполный, но симметричный, ортогональный план, насыщенный для линейной регрессионной модели.

Неполные планы при условии их **симметричности**, нормировки и **ортогональности**, позволяющие определить регрессионную модель **частного вида**, называются дробными планами или дробными репликами. Следуя этой терминологии, план табл. 26 называют *полуреplikой* полного 3-факторного плана.

Если полный план k-факторного h-уровневого эксперимента принято обозначать h^k (по его объему), то полуреplikу обозначают h^{k-1} . Указанное выражение задает объем эксперимента из N опытов. Таким образом, табл. 26 дает нам полуреplikу типа 2^{3-1} из полного плана 2^3 табл. 27. План табл. 28 можно рассматривать, как 1/16 реплика 2^{7-4} от полного плана 7-факторного эксперимента 2^7 .

Полный план k-факторного двухуровневого эксперимента можно интерпретировать, как полный перебор всех вершин k-мерного куба, полуреplikу – как перебор только половины вершин этого куба, выбранных таким образом, чтобы составить симплекс.

Как было показано выше, дробные планы можно строить из полных планов более низкого порядка. Однако при этом следует специально следить за со-

блюдением условия ортогональности. Дело в том, что ортогональность плана обеспечивает получение коэффициентов регрессии отдельно по всем факторам, в то время как неортогональные планы позволяют определить не все коэффициенты, а некоторые из них лишь для комбинаций факторов, например, вида $(x_1+x_2x_4)$ или $(x_2-5,7x_1x_2x_3)$, что не всегда удобно и приемлемо. Иными словами, ортогональный план позволяет получить систему линейных алгебраических уравнений для коэффициентов линейной регрессии, имеющих единственное решение. Т.е. *ортогональный план* обеспечивает максимум информации при заданном объеме эксперимента.

С целью соблюдения ортогональности дробные планы следует строить с помощью добавления (если это необходимо) столбцов взаимодействия. Так, для построения полуреплики 2^{4-1} 4-факторного эксперимента можно взять полный план 3-факторного – 2^3 (см. **собственно** план табл. 28) и добавить 1 столбец любого взаимодействия, например, $x_4 = x_1x_2$. Как мы видели, такая конструкция обеспечивает ортогональность, т.е. возможность определения всех коэффициентов регрессии. Эта полуреплика изображена верхними 8 строками табл. 29. В нижних 8 строках этой таблицы для наглядности помещена другая полуреплика, у которой $x_4 = -x_1x_2$. Вместе эти две полуреплики составляют *полный план* 2^4 .

Для построения четверть-реплики 5-факторного эксперимента 2^{5-2} за основу можно взять тот же полный план 2^3 , к которому следует добавить 2 столбца, составленных из **разных** произведений основных факторов.

До сих пор мы рассматривали регрессионные модели, в которых каждый фактор встречается (даже в комбинации с другими) только в степени 0 или 1. Естественно, более завершенной выглядит модель, хоть и без тройного взаимодействия, но учитывающая **степени** факторов, например, квадраты:

$$y = \lambda_0 + \sum_{j=1}^k \lambda_j x_j + \sum_{i=1}^k \sum_{j=1}^k \lambda_{ji} x_j x_i.$$

Из симметрии матрицы (λ_{ji}) следует, что для такой модели с тремя факторами $k = 3$ необходимо определить 10 различных коэффициентов. Полный же план 3-факторного эксперимента дает только $2^3 = 8$ опытов, которые не позволят их все определить. Этот *полный план для такой регрессионной модели* является *сверхнасыщенным*. Попытки расширения плана, аналогичные рассмотренным

Таблица 29.

| № опыта (i) | x_0 | x_1 | x_2 | x_3 | x_4 |
|-------------|-------|-------|-------|-------|-------|
| 1 | +1 | +1 | +1 | +1 | +1 |
| 2 | +1 | -1 | +1 | +1 | -1 |
| 3 | +1 | +1 | -1 | +1 | -1 |
| 4 | +1 | -1 | -1 | +1 | +1 |
| 5 | +1 | +1 | +1 | -1 | +1 |
| 6 | +1 | -1 | +1 | -1 | -1 |
| 7 | +1 | +1 | -1 | -1 | -1 |
| 8 | +1 | -1 | -1 | -1 | +1 |
| 9 | +1 | +1 | +1 | +1 | -1 |
| 10 | +1 | -1 | +1 | +1 | +1 |
| 11 | +1 | +1 | -1 | +1 | +1 |
| 12 | +1 | -1 | -1 | +1 | -1 |
| 13 | +1 | +1 | +1 | -1 | -1 |
| 14 | +1 | -1 | +1 | -1 | +1 |
| 15 | +1 | +1 | -1 | -1 | +1 |
| 16 | +1 | -1 | -1 | -1 | -1 |

выше, приводят к необходимости построения *неортогонального* плана, так как придется включить одинаковые столбцы с квадратами факторов: $x_j^2 = x_i^2$. А неортогональный план, как мы уже знаем, не позволит определить все коэффициенты регрессии.

Прием построения специфического неортогонального плана, позволяющего определить все коэффициенты для сложных регрессионных моделей, предложил Бокс. Следуя ему, назовем "ядром" полный план $2^3 = 8$, расширенный столбцами соответствующих взаимодействий и квадратов факторов. В табл. 30 ядро располагается в верхних 8 строках.

Таблица 30.

| № опыта | x_0 | x_1 | x_2 | x_3 | x_1^2 | x_2^2 | x_3^2 | x_1x_2 | x_1x_3 | x_2x_3 | |
|---------|-------|-----------|-----------|-----------|------------|------------|------------|----------|----------|----------|--------------|
| 1 | +1 | +1 | +1 | +1 | +1 | +1 | +1 | +1 | +1 | +1 | я |
| 2 | +1 | -1 | +1 | +1 | +1 | +1 | +1 | -1 | -1 | +1 | |
| 3 | +1 | +1 | -1 | +1 | +1 | +1 | +1 | -1 | +1 | -1 | д |
| 4 | +1 | -1 | -1 | +1 | +1 | +1 | +1 | +1 | -1 | -1 | |
| 5 | +1 | +1 | +1 | -1 | +1 | +1 | +1 | +1 | -1 | -1 | р |
| 6 | +1 | -1 | +1 | -1 | +1 | +1 | +1 | -1 | +1 | -1 | |
| 7 | +1 | +1 | -1 | -1 | +1 | +1 | +1 | -1 | -1 | +1 | о |
| 8 | +1 | -1 | -1 | -1 | +1 | +1 | +1 | +1 | +1 | +1 | |
| 9 | +1 | $+\alpha$ | 0 | 0 | α^2 | 0 | 0 | 0 | 0 | 0 | з |
| 10 | +1 | $-\alpha$ | 0 | 0 | α^2 | 0 | 0 | 0 | 0 | 0 | в |
| 11 | +1 | 0 | $+\alpha$ | 0 | 0 | α^2 | 0 | 0 | 0 | 0 | е |
| 12 | +1 | 0 | $-\alpha$ | 0 | 0 | α^2 | 0 | 0 | 0 | 0 | з |
| 13 | +1 | 0 | 0 | $+\alpha$ | 0 | 0 | α^2 | 0 | 0 | 0 | д |
| 14 | +1 | 0 | 0 | $-\alpha$ | 0 | 0 | α^2 | 0 | 0 | 0 | ы |
| 15 | +1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | центр |

Дополнять план Бокс предложил, искусственно **увеличивая количество уровней** факторов введением в план новых точек на ребрах k -мерного куба. Необходимость этого ясна, если вспомнить, что выявить квадратичную зависимость можно только по не менее чем трем точкам, т.е. необходимо иметь не менее трех уровней факторов. $2k$ "звездных" точек представляют собой точки плана с координатами по основным факторам вида $(\pm\alpha, 0, 0, \dots, 0)$, $(0, \pm\alpha, 0, 0, \dots, 0)$, \dots , $(0, 0, \dots, \pm\alpha)$ – они расположены в табл. 30 в строках с 9-ой по 14-ю. Еще одна точка – центральная – имеет все координаты $x_j = 0$, она помещена в 15-ой строке. Таким образом, план стал **ненасыщенным**, что позволяет кроме всех требуемых коэффициентов регрессии провести оценку некоторых статистических критериев или дисперсионный анализ за счет оставшихся $(15 - 10) = 5$ сте-

пений свободы. Заметим, кстати, что во всех формулах вычисления дисперсии при таком анализе число степеней свободы должно быть **уменьшено** на количество найденных коэффициентов регрессии (в нашем примере на 10).

7.8. Сравнение приемов планирования эксперимента

Для подробного изучения нескольких возможных **приемов** целенаправленного построения планов рассмотрим следующий пример планирования эксперимента для дисперсионного анализа.

Для некоторого технологического процесса требуется оценить эффекты обработки 5 различными способами: А, В, С, D, Е (фактор обработки). Поставим задачу **сравнить возможные планы** эксперимента, состоящего в распределении способов обработки по объектам, с точки зрения наилучшей точности и наибольшей информативности.

В качестве исходного экспериментального материала предоставлены 35 объектов производства, подлежащих дальнейшей обработке исследуемыми способами. Результаты контрольного замера в этой партии объектов до обработки представлены в 4-м столбце табл. 31. Каждое значение в этом столбце является суммой двух **неподдающихся измерениям** значений: истинного значения измеряемого параметра (2-й столбец) и погрешности замера (3-й столбец), которая предполагается нормально распределенной. Таким образом, например, двадцатый объект поступает на обработку одним из исследуемых способов с параметром: $26,7 = 24 + 2,7$, который тоже **не поддается замеру**.

Эффекты каждого способа обработки постоянны, не зависят от исходных параметров объекта и выражаются в добавлении определенного значения к контролируемому параметру: для А – это $(+3,0)$, для В – $(+1,5)$, для С – $(0,0)$, для D – $(-1,5)$, для Е – $(-3,0)$.

Заметим, что при последующем проведении планируемого эксперимента ни эти значения, которые и следует оценить, ни данные 4-го столбца (параметр до обработки) **известны не будут** – контролируемым останется только результат обработки, например, двадцатого объекта способом Е: $23,7 = 26,7 - 3,0$. Т.е. влияние каждого способа обработки придется оценивать лишь по этим замерам. В табл. 31 исходный статистический материал в столбцах 2, 3, 4, характеризующий предварительные стадии производственного процесса, дан лишь для наглядности сравнения различных планов эксперимента.

а) Рассмотрим первый пример – план **упорядоченного** (нерандомизированного) распределения способов обработки, представленный 5-м столбцом табл. 31. Проанализируем результаты этого эксперимента, т.е. поддающиеся контролю результаты замера параметра после обработки. Эти данные показаны в табл. 32, сгруппированными по способам обработки, под которые попало ровно по 7 объектов.

| №
объ-
екта | Истинное
значение
параметра | Погреш-
ность
замера | Результат
замера
параметра
до
обработки | Планы экспериментов | | | | |
|-------------------|-----------------------------------|----------------------------|---|--|-------------------------------------|---|---|--|
| | | | | Упорядо-
ченное
распре-
деление | Полностью
рандомизи-
рованный | Рандоми-
зирован-
ные пол-
ные блоки | Латинский
квадрат
(полные
блоки) | Система-
тическое
распре-
деление |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 9 | -1,4 | 7,6 | A | D | B | C | D |
| 2 | 12 | 2,8 | 14,8 | A | D | E | E | E |
| 3 | 11 | 3,6 | 14,6 | A | B | A | A | A |
| 4 | 12 | -1,7 | 10,3 | A | E | D | D | B |
| 5 | 16 | 0,4 | 16,4 | A | C | C | B | C |
| 6 | 17 | 2,9 | 19,9 | A | D | E | B | D |
| 7 | 14 | 0,4 | 14,4 | A | B | C | D | E |
| 8 | 14 | 1,4 | 15,4 | B | A | A | E | A |
| 9 | 17 | -2,0 | 15,0 | B | C | D | C | B |
| 10 | 16 | 1,6 | 17,6 | B | B | B | A | C |
| 11 | 13 | 0,0 | 13,0 | B | C | B | A | D |
| 12 | 12 | -2,3 | 9,7 | B | A | A | C | E |
| 13 | 12 | -3,1 | 8,9 | B | C | D | D | A |
| 14 | 17 | -0,1 | 16,9 | B | A | C | B | B |
| 15 | 20 | -0,4 | 19,6 | C | D | E | E | C |
| 16 | 20 | -1,0 | 19,0 | C | A | B | E | D |
| 17 | 20 | -3,2 | 16,8 | C | E | C | B | E |
| 18 | 22 | 1,6 | 23,6 | C | D | A | C | A |
| 19 | 23 | 1,4 | 24,4 | C | B | E | A | B |
| 20 | 24 | 2,7 | 26,7 | C | B | D | D | C |
| 21 | 23 | 1,8 | 24,8 | C | E | C | D | D |
| 22 | 26 | 1,5 | 27,5 | D | B | E | A | E |
| 23 | 26 | -1,7 | 24,3 | D | A | B | B | A |
| 24 | 27 | 0,5 | 27,5 | D | A | D | E | B |
| 25 | 26 | 0,6 | 26,6 | D | D | A | C | C |
| 26 | 25 | 0,3 | 25,3 | D | C | D | | D |
| 27 | 26 | -0,4 | 25,6 | D | C | B | | E |
| 28 | 29 | 3,5 | 32,5 | D | B | C | | A |
| 29 | 29 | 0,4 | 29,4 | E | C | E | | B |
| 30 | 28 | 2,5 | 30,5 | E | E | A | | C |
| 31 | 30 | -1,2 | 28,8 | E | E | D | | D |
| 32 | 31 | 0,3 | 31,3 | E | A | A | | E |
| 33 | 30 | -2,7 | 27,3 | E | E | B | | A |
| 34 | 30 | -3,8 | 26,2 | E | D | E | | B |
| 35 | 28 | 4,5 | 32,5 | E | E | C | | |
| | 21,00
46,18 | 0,28
4,404 | 21,28
52,49 | Среднее
Дисперсия | | | | |

| №
объекта (i) | Способы обработки (j) | | | | |
|-----------------------------|-----------------------|--------|-------|-------|-------|
| | A | B | C | D | E |
| 1 | 10.6 | 16.9 | 19.6 | 26.0 | 26.4 |
| 2 | 17.8 | 16.5 | 19.0 | 22.8 | 27.5 |
| 3 | 17.6 | 19.1 | 16.8 | 26.0 | 25.8 |
| 4 | 13.3 | 14.5 | 23.6 | 25.1 | 28.3 |
| 5 | 19.4 | 11.2 | 24.4 | 23.8 | 24.3 |
| 6 | 22.9 | 10.4 | 26.7 | 24.1 | 23.2 |
| 7 | 17.4 | 18.4 | 24.8 | 31.0 | 29.5 |
| \bar{Y}_j | 17.00 | 15.29 | 22.13 | 25.54 | 26.43 |
| $\bar{Y}_j - \bar{\bar{Y}}$ | - 4.28 | - 5.99 | 0.85 | 4.26 | 5.15 |
| $\bar{\bar{Y}} = 21.28$ | | | | | |

Здесь: \bar{Y}_j – среднее значение контролируемого параметра изделий в группе, обработанной j-ым способом, $\bar{\bar{Y}}$ – среднее значение параметра всех объектов после обработки. Нетрудно видеть, что величина $\bar{Y}_j - \bar{\bar{Y}}$ характеризует **искомую оценку** эффекта способов обработки, поскольку задает отклонение средней величины измеряемого фактора в группе одного из способов обработки от общего среднего.

Первые 7 объектов, прошедших обработку способом А дали среднее значение контролируемого параметра 17,00; вторые при способе В – 15,29; далее соответственно 22,13; 25,54; 26,43. Создается впечатление, что эффекты первых способов обработки меньше, чем последних, что ошибочно! Таким образом, план упорядоченного распределения способов обработки может приводить к выводам, прямо противоположным истинному положению вещей. Это говорит о том, что на итоговый наблюдаемый результат замеров оказывают существенное влияние неизвестные систематические отклонения, не связанные с применяемым способом обработки. При таком плане эксперимента систематические отклонения исходного материала **затмевают** результаты обработки. Можно сказать, что эксперимент неудачно спланирован.

Таким образом, можно сделать вывод, что результаты упорядоченного *плана однофакторного эксперимента* весьма чувствительны к скрытым факторам исходного статистического материала.

б) Для получения менее безнадёжных результатов достаточно распределить способы обработки по объектам так, чтобы уравнивать возможности появления различных отклонений в исходном материале, т.е. достаточно применить *рандомизацию*. В 6-ом столбце табл. 31 все 5 способов обработки распределены между объектами случайным образом, т.е. каждый объект подвергается какому-либо способу обработки **независимо** от отклонения контролируемого параметра и порядкового номера.

В верхней части табл. 33 представлены результаты **полностью рандомизированного** эксперимента, сгруппированные по способам обработки. Наблюдаемое значение параметра, например, восьмого объекта, подвергнувшегося согласно плану обработке способом А, 18,4 – складывается из значения параметра до обработки 15,4 и результата обработки +3,0 способом А. В нижней части табл. 33 приводятся все данные дисперсионного анализа влияния способов обработки для оценки их вклада.

| №
объекта | Способы обработки | | | | |
|--------------------------------------|-------------------|-------|-------|-------|-------|
| | A | B | C | D | E |
| 1 | 18,4 | 16,1 | 16,4 | 6,1 | 7,3 |
| 2 | 12,7 | 15,9 | 15,0 | 13,3 | 13,8 |
| 3 | 19,9 | 19,1 | 13,0 | 18,4 | 21,8 |
| 4 | 22,0 | 25,9 | 8,9 | 18,1 | 27,5 |
| 5 | 27,3 | 28,2 | 25,3 | 22,1 | 25,8 |
| 6 | 30,5 | 29,0 | 25,6 | 25,1 | 24,3 |
| 7 | 34,3 | 34,0 | 29,4 | 24,7 | 29,5 |
| \bar{y}_j | 23,59 | 24,03 | 19,09 | 18,26 | 21,43 |
| $\bar{y}_j - \bar{\bar{y}}$ | 2,31 | 2,75 | -2,19 | -3,02 | 0,15 |
| $\bar{\bar{y}} = 21,28$ | | | | | |
| Дисперсионный анализ | | | | | |
| Оценка рассеяния | | | | f | s^2 |
| между группами по способам обработки | | | | 4 | 46,98 |
| внутренняя | | | | 30 | 55,04 |

Здесь: f – число степеней свободы, s^2 – статистическая оценка дисперсий: **между группами** по способам обработки:

$$s_A^2 = \frac{\sum_{j=1}^k N_j (\bar{y}_j - \bar{\bar{y}})^2}{k - 1},$$

где $k = 5$ по числу способов обработки, $N_j = 7$ по числу объектов, обработанных j -ым способом; и остаточной – **внутренней**:

$$s_0^2 = \frac{\sum_{i=1}^m \sum_{j=1}^k (y_{ij} - \bar{y}_j)^2}{\sum_{j=1}^k N_j - k},$$

где y_{ij} – значение параметра i -го объекта из обработанных j -ым способом, число объектов, обработанных каждым способом, $m = 7$.

Внутренняя дисперсия дает среднее квадратическое отклонение: $s_0 = \sqrt{55,04 / 7} = 2,80$. Это значение сравнимо с самой величиной $\bar{y}_j - \bar{\bar{y}}$ и слишком велико, чтобы сделать оценку эффекта способов обработки уверенной. Это и проявилось в весьма далеких от истины оценках, полученных при этом рандомизированном *плане однофакторного эксперимента*.

Таким образом, можно сделать вывод, что рандомизация плана позволяет существенно улучшить информативность эксперимента, но влияние скрытых факторов не выявляет.

в) Уменьшению внутренней дисперсии могли бы способствовать два мероприятия: увеличение знаменателя за счет повторения каждого способа обработки более 7 раз (увеличения N_j) или уменьшение числителя, т.е. дисперсии ошибки. Первый требует увеличения объема эксперимента, что не всегда возможно. Второй, как теперь видно, требует более изощренного построения плана эксперимента. Из основного уравнения дисперсионного анализа (§ 6.3) следует, что общая дисперсионная сумма разбивается на столько частей, сколько факторов исследуется в задаче. Следовательно, надо **увеличивать** количество учитываемых факторов и, соответственно, частных дисперсий.

В нашем экспериментальном материале разнородность определяется неучтенным влиянием неизвестных факторов (например, местом или временем изготовления) – из последней строки табл. 31 видно, что они привели к сильному разбросу исходного истинного значения измеряемого параметра, дисперсия которого 46,18 во много раз больше дисперсии погрешности замера 4,404. Разделим весь экспериментальный материал на одинаковые **блоки** по порядку. Порядок блоков и будем считать вторым (искусственным) фактором для построения *плана двухфакторного эксперимента*. Для того чтобы в каждом из этих блоков можно было бы применить обработку всеми оцениваемыми способами, их объем должен быть равен числу способов обработки $k = 7$.

Ясно, что с целью рандомизации, (как в предыдущем варианте б) т.е. случайного распределения оцениваемых способов обработки, их следует применить внутри каждого блока к случайным объектам. Так составлен план эксперимента с помощью **рандомизированных полных блоков** (7-ой столбец табл. 31). Табл. 34 построена аналогично табл. 33 и дает возможность провести дисперсионный анализ.

Таблица 34.

| № блока | Способы обработки | | | | | \bar{Y}_i |
|--------------------------------------|-------------------|-------|-------|-------|-------|-------------|
| | A | B | C | D | E | |
| 1 | 17,6 | 9,1 | 16,4 | 8,8 | 11,8 | 12,74 |
| 2 | 18,4 | 19,1 | 14,4 | 13,5 | 16,9 | 16,46 |
| 3 | 12,7 | 14,5 | 16,9 | 7,4 | 16,6 | 13,62 |
| 4 | 26,6 | 20,5 | 16,8 | 25,2 | 21,4 | 22,10 |
| 5 | 29,6 | 25,8 | 24,8 | 26,0 | 24,5 | 26,14 |
| 6 | 33,5 | 27,1 | 32,5 | 23,8 | 26,4 | 28,66 |
| 7 | 34,3 | 28,8 | 32,5 | 27,3 | 23,2 | 29,22 |
| \bar{Y}_j | 24,67 | 20,70 | 22,04 | 18,86 | 20,11 | |
| $\bar{y}_j - \bar{\bar{y}}$ | 3,39 | -0,58 | 0,76 | -2,42 | -1,17 | |
| $\bar{\bar{y}} = 21,28$ | | | | | | |
| Дисперсионный анализ | | | | | | |
| Оценка рассеяния | | | | | f | s^2 |
| между блоками | | | | | 6 | 247,20 |
| между группами по способам обработки | | | | | 4 | 34,39 |
| внутренняя | | | | | 24 | 9,858 |

Здесь статистические оценки частных дисперсий S^2 подсчитываются следующим образом: **между блоками** по формуле:

$$s_1^2 = \frac{\sum_{i=1}^b m_i (\bar{y}_i - \bar{\bar{y}})^2}{b-1},$$

где $b = 7$ по числу блоков, $m_j = 5$ по числу объектов в j -м блоке; **между группами** по способам обработки – по формуле:

$$s_A^2 = \frac{\sum_{j=1}^k N_j (\bar{y}_j - \bar{\bar{y}})^2}{k-1};$$

а остаточная, **внутренняя** – по формуле:

$$s_0^2 = \frac{\sum_{i=1}^b \sum_{j=1}^k (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{\bar{y}})^2}{(b-1)(k-1)}.$$

В этом случае остаточное среднее квадратическое отклонение $s_0 = \sqrt{9,858/7} = 1,19$ существенно меньше, чем в полностью рандомизированном эксперименте. Поэтому оценка эффекта способов обработки $\bar{y}_j - \bar{\bar{y}}$ существенно лучше и ближе к истине.

Таким образом, можно сделать два вывода. Во-первых, чем меньше остаточная дисперсия, тем ближе среднее выборочное значение исследуемого фактора к его математическому ожиданию, т.е. к истине. Во-вторых, для уменьшения остаточной дисперсии следует вводить как можно больше дополнительных факторов. На последний вывод следует обратить особое внимание, так как он требует совершенно иную постановку эксперимента на больших и "плохо организованных системах", чем на "хорошо организованных системах". А именно: для получения добротных результатов статистической обработки данных эксперимента следует учитывать случайным образом **как можно большее число факторов!**

г) Построим план эксперимента по трем факторам – план трехфакторного эксперимента. Введем в рассмотрение в качестве третьего, искусственного фактора порядок способов обработки внутри блока. Блоки по-прежнему возьмем размером $m = 5$ по количеству способов обработки (полные блоки). Убедившись в благотворности рандомизации, обеспечим ее с помощью случайного выбора порядка в блоке.

Рандомизацию и сбалансированность блоков могут обеспечить латинские квадраты, в которых порядок расположения всех пяти способов и блоков уникален (§ 7.2). Таким образом, рандомизация распространена как на блоки, так и на порядки. Случайно выбранный латинский квадрат 5×5 приведен в табл. 35 (он составлен по упоминавшемуся правилу по столбцам с дополнительной перестановкой столбцов местами). В 8-ом столбце табл. 31 показан план, составленный из блоков – столбцов этого латинского квадрата табл. 35.

Таблица 35.

| | | | | |
|---|---|---|---|---|
| С | Е | А | Д | В |
| В | Д | Е | С | А |
| А | С | Д | В | Е |
| Е | В | С | А | Д |
| Д | А | В | Е | С |

Но таких блоков по 5 элементов на нашем экспериментальном материале может быть 7, а в латинском квадрате их только 5. Очевидно, часть статистического материала не будет

использоваться. Тем более интересно, насколько добротный результат может быть получен в этом случае на меньшем объеме эксперимента.

В табл. 36 приведен дисперсионный анализ результатов эксперимента, спланированного с помощью латинских квадратов.

Таблица 36.

| № блока | Порядок внутри блока | | | | | \bar{Y}_i |
|--------------------------------------|----------------------|--------|--------|--------|--------|-------------|
| | 1 | 2 | 3 | 4 | 5 | |
| 1 | С 7,6 | Е 11,8 | А 17,6 | Д 8,8 | В 17,9 | 12,74 |
| 2 | В 21,4 | Д 12,9 | Е 12,4 | С 15,0 | А 20,6 | 16,46 |
| 3 | А 16,0 | С 9,7 | Д 7,4 | В 18,4 | Е 16,6 | 13,62 |
| 4 | Е 16,0 | В 18,3 | С 23,6 | А 27,4 | Д 25,2 | 22,10 |
| 5 | Д 23,3 | А 30,5 | В 25,8 | Е 24,5 | С 26,6 | 26,14 |
| \bar{Y}_s | 16,86 | 16,64 | 17,36 | 18,82 | 21,38 | |
| $\bar{y} = 18,21$ | | | | | | |
| Способы обработки | | | | | | |
| | А | В | С | Д | Е | |
| \bar{y}_j | 22,42 | 20,36 | 16,50 | 15,52 | 16,26 | |
| $\bar{y}_j - \bar{y}$ | 4,21 | 2,15 | -1,71 | -2,69 | -1,95 | |
| Дисперсионный анализ | | | | | | |
| Оценка рассеяния | | | | f | s^2 | |
| между блоками | | | | 4 | 165,09 | |
| между порядками | | | | 4 | 19,29 | |
| между группами по способам обработки | | | | 4 | 45,39 | |
| внутренняя | | | | 12 | 10,03 | |

При этом частные дисперсии **между блоками** и **между порядками** рассчитываются по формулам:

$$s_2^2 = \frac{\sum_{i=1}^b m_i (\bar{y}_i - \bar{y})^2}{b-1} \quad \text{и} \quad s_1^2 = \frac{\sum_{s=1}^k l_s (\bar{y}_s - \bar{y})^2}{k-1},$$

соответственно, в которых $k = b = l_s = m_j = 5$. Частная дисперсия **между группами** по способам обработки задается формулой:

$$s_A^2 = \frac{\sum_{j=1}^k N_j (\bar{y}_j - \bar{y})^2}{k-1},$$

где $N_j = 5$, а остаточная, **внутренняя** дисперсия:

$$s_0^2 = \frac{\sum_{i=1}^b \sum_{s=1}^k (y_{sj} - \bar{y}_i - \bar{y}_j - \bar{y}_s + 2\bar{\bar{y}})^2}{(k-1)(b-2)},$$

где j принимает единственное значение, соответствующее единственному способу обработки при выбранных s и i .

Напомним, что все эти частные дисперсии в соответствии с основным уравнением дисперсионного анализа (§ 6.3) с учетом степеней свободы дают в сумме общую дисперсию:

$$s^2 = \frac{\sum_{i=1}^b \sum_{j=1}^k (y_{ij} - \bar{\bar{y}})^2}{(k-1)(b-1)}.$$

Прежде чем перейти к оценке результатов планирования эксперимента с помощью латинских квадратов, заметим, что в последнем плане требуемый объем экспериментального материала значительно **меньше** предыдущих (25 объектов вместо 35).

Как показывает дисперсионный анализ табл. 36, рассеяние между порядками внутри блоков несущественно по сравнению с рассеянием между блоками и способами обработки, т.е. двойной классификацией (введением второго дополнительного оцениваемого фактора) заметного уменьшения внутренней дисперсии мы не добились. Это значит, что такого дополнительного фактора, скорее всего, не существует. (Между прочим, критерий Фишера § 6.3 позволяет оценить значимость таких оценок рассеяния.) Однако на существенно меньшем объеме статистического материала удалось как бы "очистить" остаточное рассеяние s_0 оценки эффекта способов обработки $\bar{y}_j - \bar{\bar{y}}$ от наносного рассеяния прочих факторов и таким образом уменьшить его до $s_0 = \sqrt{10,03/5} = 1,42$. Нетрудно заметить, что полученные данным способом оценки эффекта методов обработки $\bar{y}_j - \bar{\bar{y}}$ не уступают случаю рандомизированных полных блоков и достаточно близки к истине.

Таким образом, можно сделать вывод, что если в предоставленном экспериментальном материале существенное значение имеют какие-то два неучтенных фактора, а не один, то их влияние на итоговую оценку эффекта обработки можно существенно снизить. Более того, если второй (третий и т.п.) дополнительный фактор несущественен, то его введение не повлияет на результаты дисперсионного анализа и выводы по результатам эксперимента.

д) Итак, рандомизация действительно позволила учесть систематическое влияние неучтенных случайных факторов и получить достаточно близкие к истине оценки исследуемого фактора. Однако если все основные факторы **учесть предварительно** и поставить под контроль, то рандомизация может и не понадобиться.

В качестве примера **нерандомизированного** плана рассмотрим план систематического распределения, представленный 9-ым столбцом табл. 31. **Используя** полученные в предыдущих исследованиях **знания**, что единственный "мешающий" фактор приводит к плавному возрастанию **по номерам** систематического отклонения контролируемого параметра, можно построить статистические оценки на основании "скользящих средних" по пяти соседним объектам. В этом случае именно небольшие отклонения между соседними объектами позволяют надеяться на такой же эффект, как и при разбиении на блоки. Однако в данном плане можно обойтись без рандомизации, так как влияние основного мешающего фактора **учтено**. Поэтому в рассматриваемом нерандомизированном плане способы обработки распределены систематически – расположены упорядочено (А, В, С, D, Е).

Таблица 37.

Скользящие средние строятся следующим образом (табл. 37). Каждые 5 смежных объектов, обработанных разными способами (потому и необходимо систематическое расположение А, В, С, D, Е), порождают для центрального объекта из этой пятерки скользящие суммы (3-ий столбец) и скользящие средние арифметические (4-ый столбец) значений контролируемого параметра после обработки. В 5-ом столбце располагаются скользящие разности d_{tj} между значениями параметра после обработки и скользящими средними – между 2-ым и 4-ым столбцами.

Способы обработки (А, В, С, D, Е) приходится располагать в указанном порядке, начиная от третьего объекта, так как именно он порождает первую пятерку данных и именно от него определяется первая скользящая разность. Неполные (крайние) блоки этой системы способов обработки составляют так, как если бы они были полными и продолжались дальше.

Результаты обработки данных такого эксперимента представлены в табл. 38. В ней скользящие разности сгруппированы по соответствующим им способам обработки, а \bar{d}_j представляет собой среднюю их величину для j-го способа, т.е. оценку эффекта обработки, которая оказалась весьма близкой к истине.

В итоге по такому плану каждому способу обработки подвергнуты $c = 6$ объектов, что позволяет исходный экспериментальный материал. Поэтому сумму квадратов отклонений можно разбить, следуя Фишеру, следующим образом:

$$\sum_{t=1}^c \sum_{j=1}^k d_{tj}^2 = \sum_{t=1}^c \sum_{j=1}^k (d_{tj} - \bar{d}_j)^2 + c \sum_{j=1}^k \bar{d}_j^2.$$

Здесь последнее слагаемое в правой части, деленное на $k - 1 = 4$, дает оценку частной дисперсии **между группами** по 5 способам обработки, а первое, деленное на $(c - 1)(k - 1) = 20$ – остаточную **внутреннюю** дисперсию искомой оценки.

Остаточное среднее квадратическое отклонение получается равным $s_0 = \sqrt{7,477/6} = 1,12$, что позволяет считать полученные оценки эффекта способов обработки достаточно точными и не подверженными каким-либо существенным неучтенным факторам.

| № объекта | Значение параметра после обработки | Скользящие суммы | Скользящие средние | Скользящие разности d_{tj} |
|-----------|------------------------------------|------------------|--------------------|------------------------------|
| 1 | 2 | 3 | 4 | 5 |
| 1 | 6,1 | 63,7 | 12,74 | 4,86 |
| 2 | 11,8 | | | |
| 3 | 17,6 | | | |
| 4 | 11,8 | | | |
| 5 | 16,4 | | | |
| 6 | 18,4 | 76,0 | 15,20 | -3,40 |
| 7 | 11,4 | 75,6 | 15,12 | 1,28 |
| 8 | 18,4 | 76,4 | 15,28 | 3,12 |
| 9 | 11,4 | 81,1 | 16,22 | -4,82 |
| 10 | 18,4 | 82,3 | 16,46 | 1,94 |
| 11 | 16,5 | 75,4 | 15,08 | 1,42 |
| 12 | 17,6 | 70,7 | 14,14 | 3,46 |
| 13 | 11,5 | 64,2 | 12,84 | -1,34 |
| 14 | 6,7 | 66,1 | 13,22 | -6,52 |
| 15 | 11,9 | 68,1 | 13,62 | -1,72 |
| 16 | 18,4 | 74,1 | 14,82 | 3,58 |
| 17 | 19,6 | 81,2 | 16,24 | 3,36 |
| 18 | 13,8 | 103,4 | 20,68 | -6,88 |
| 19 | 26,6 | 110,5 | 22,10 | 4,50 |
| 20 | 25,9 | 116,3 | 23,26 | 2,64 |
| 21 | 26,7 | 127,0 | 25,40 | 1,30 |
| 22 | 23,3 | 127,7 | 25,54 | -2,24 |
| 23 | 24,5 | 130,8 | 26,16 | -1,66 |
| 24 | 27,3 | 130,7 | 26,14 | 1,16 |
| 25 | 29,0 | 131,2 | 26,24 | 2,76 |
| 26 | 26,6 | 129,3 | 25,86 | 0,74 |
| 27 | 23,8 | 137,5 | 27,50 | -3,70 |
| 28 | 22,6 | 139,4 | 27,88 | -5,28 |
| 29 | 35,5 | 143,3 | 28,66 | 6,84 |
| 30 | 30,9 | 146,8 | 29,36 | 1,54 |
| 31 | 30,5 | 152,5 | 30,50 | 0,00 |
| 32 | 27,3 | 147,3 | 29,46 | -2,16 |
| 33 | 28,3 | 144,1 | 28,82 | -0,52 |
| 34 | 30,3 | | | |
| 35 | 27,7 | | | |

| № блока | Способы обработки | | | | |
|--------------------------------------|-------------------|-------|------|-------|----------------|
| | A | B | C | D | E |
| 1 | 4,86 | -3,40 | 1,28 | 3,12 | -4,82 |
| 2 | 1,94 | 1,42 | 3,46 | -1,34 | -6,52 |
| 3 | -1,72 | 3,58 | 3,36 | -1,68 | -6,88 |
| 4 | 4,50 | 2,64 | 1,30 | -2,24 | -1,66 |
| 5 | 1,16 | 2,76 | 0,74 | -3,70 | -5,28 |
| 6 | 6,84 | 1,54 | 0,00 | -2,16 | -0,52 |
| \bar{d}_j | 2,93 | 1,42 | 1,69 | -1,33 | -4,28 |
| Дисперсионный анализ | | | | | |
| Оценка рассеяния средних разностей | | | | f | s ² |
| между группами по способам обработки | | | | 4 | 50,34 |
| внутренняя | | | | 20 | 7,477 |

Рассмотренный пример демонстрирует положительные и отрицательные стороны различных планов эксперимента для дисперсионного анализа. Оказывается возможным **выявить эффект воздействия частного фактора**, наличие которого лишь предполагается. Эта возможность основывается на дисперсионном анализе, позволяющем селективировать систематические отклонения (воздействия) от случайных посредством разбиения общей дисперсии.

Таким образом, в арсенале средств планирования эксперимента оказываются следующие приемы, повышающие качество параметрической оценки эффекта посредством уменьшения остаточной дисперсии:

- рандомизация,
- искусственное увеличение количества учитываемых факторов,
- разбиение на блоки и группы,
- латинские квадраты.

В заключение данного параграфа заметим, что все рассмотренные в примере *планы сбалансированы* по способам обработки – исследуемому фактору.

Глава 8. Особые методы планирования эксперимента

8.1. Специальные приемы планирования эксперимента

Классическая теория планирования эксперимента, изложенная в главе 7, как мы видели, представляет собой набор приемов составления матрицы дискретного эксперимента с конечным числом уровней факторов. В этом смысле планирование эксперимента с непрерывными значениями факторов принципиально невозможно и приходится искусственно вводить **дискретные** уровни. Если рассматриваемые факторы имеют вид простых измеримых параметров, то проблема состоит лишь в том, чтобы выбрать наименьшее число наиболее представительных значений в качестве уровней. Куда более сложным представляется случай неизмеримых, качественных факторов. Для их дискретизации необходимы неформальные процедуры, основанные на глубоком знании особенностей объекта. Приходится применять методы *ранжирования* и *экспертных оценок* (§ 3.1 и § 8.2).

1) Для многофакторных экспериментов большое число уровней может стать непреодолимым препятствием не только при планировании, но и при постановке отбора и обработке данных. Само по себе большое число факторов увеличивает лишь размерность задачи и усложняет ее трактовку. Хуже обстоит дело при многоуровневых факторах – план и эксперимент в этом случае теряют наглядность. Поэтому на практике стараются, прежде всего, уменьшить именно число уровней.

Конечно, это не является самоцелью. Объем эксперимента определяется произведением: $\prod_{j=1}^k h_j$, где k – число факторов, а h_j – число уровней каждого фактора. Если все факторы имеют одинаковое число уровней $h = h_j$, то объем эксперимента равен h^k . Проанализируем это выражение. При $h = 2$ увеличение k ведет к более слабому росту объема эксперимента, чем увеличение h . Поэтому при необходимости расширения двухуровневого эксперимента следует по возможности сохранить минимальное число уровней, может быть даже за счет введения новых факторов. При $h = 3$ такая же ситуация наблюдается для $k \geq 4$, а при $h = 4$ – для $k \geq 7$. Т.е. при меньшем, чем указано, количестве факторов выгоднее увеличивать на единицу число уровней, а не факторов. Вообще говоря, в каждом конкретном случае следует выбирать оптимальные значения k и h_j , исходя из особенностей объекта, а не только для минимизации объема эксперимента.

2) Еще одной серьезной проблемой при планировании эксперимента является выполнение требований дисперсионного анализа, особенно в части **независимости** исследуемых факторов (§ 6.3). На практике изучение сложных объектов обычно происходит в режиме пассивного наблюдения. При этом выбор изучаемых факторов часто бывает ограничен только теми, которые легко регистрировать. В этих условиях априори гарантировать независимость наблюдаемых параметров невозможно.

Решению этой проблемы помогает метод главных компонент, который позволяет выбрать исследуемые факторы на основе их **некоррелированности**. Поскольку для дисперсионного анализа непосредственное значение имеет именно отсутствие корреляции между исследуемыми параметрами, то последний факт можно использовать вместо громоздких и неформальных исследований их независимости. В конечном итоге можно даже удовлетвориться ненулевой, но относительно **слабой корреляцией**, оценку которой следует делать для каждой конкретной задачи отдельно.

Основная идея метода главных компонент состоит в замене переменных, характеризующих исследуемые факторы. Пусть исходная система из k наблюдаемых входных факторов x_j ($j = 1, 2, \dots, k$) построена по итогам N наблюдений и представляется матрицей результатов эксперимента:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nk} \end{pmatrix},$$

в которой столбцами являются наблюдаемые значения входных факторов. (Не трудно заметить тождественность этой матрицы с планом эксперимента – это и есть план, но пассивного эксперимента.) По этим данным можно рассчитать матрицу выборочных *ковариаций* (ковариационную матрицу), элементами которой являются:

$$l_{ij} = \frac{1}{N-1} \cdot \sum_{n=1}^N (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j),$$

и корреляционную матрицу $R = (r_{ij})$, составленную из выборочных оценок *коэффициентов корреляции*:

$$r_{ij} = \frac{1}{s_i s_j} \cdot \frac{1}{N-1} \cdot \sum_{n=1}^N (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j),$$

где s_i – выборочная оценка среднего квадратического отклонения i -го фактора. Заметим, что диагональные элементы матрицы R равны единице, а сама матрица – симметричная:

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & 1 \end{pmatrix}.$$

Замена переменных

$$z_i = \sum_{j=1}^k a_{ij} x_j$$

подбирается таким образом, чтобы новые переменные z_i были некоррелированы. Для этого достаточно, чтобы матрица ковариаций новых переменных стала диагональной. Для того, чтобы выбор шкал измерения факторов не привносил в расчеты своих масштабов, следует максимально унифицировать системы отсчета и нормировать переменные.

Математический метод решения такой задачи известен и реализуется с помощью стандартного программного обеспечения современных ЭВМ. Из уравнения $\det(R - vE) = 0$, где E – единичная матрица, находятся собственные значения матрицы R : $v_i = s^2(z_i)$, которые позволяют вычислить собственные векторы \mathbf{a}_i той же матрицы из матричных уравнений $R\mathbf{a}_i = v_i\mathbf{a}_i$. В свою очередь собственные векторы определяют матрицу искомого преобразования, как матрицу, составленную из векторов \mathbf{a}_i как столбцов: $A = (\mathbf{a}_{mi})$.

В процессе указанных математических преобразований могут встретиться затруднения в виде невозможности получения невырожденных решений некоторых матричных уравнений. Это будет свидетельствовать о неполноте исходного перечня факторов, их неустранимой зависимости или коррелированности. В этом случае представляются возможными два пути. Во-первых, расширение эксперимента и введение в рассмотрение дополнительных исходных факторов, из которых затем будут получены новые некоррелированные. А во-вторых, без изменения исходной базы факторов подбор новых переменных со слабой (минимальной) корреляцией между собой. В этом случае приходится решать задачу оптимизации с громоздким критерием и множеством переменных. Однако, принципиально такая задача разрешима.

После получения, таким образом, новых некоррелированных переменных дисперсионный анализ может быть проведен без каких-либо ограничений. По тем нескольким переменным, у которых дисперсия наибольшая, можно провести построение регрессионной модели. Но при применении этой модели могут встретиться трудности интерпретации новых переменных – некоррелированных факторов, которые определяются с помощью некоторых линейных комбинаций исходных. Иногда для обеспечения научно-технологического прорыва приходится ломать привычные стереотипы образного представления об изучаемом явлении и вводить новые понятия и факторы.

3) При планировании эксперимента немаловажное значение имеет удачный выбор факторов, который позволил бы уменьшить его объем. Задачу выбора наименьшего числа наиболее представительных (т.е., как мы теперь знаем, независимых – некоррелированных) факторов помогает решить факторный анализ. Если в методе главных компонент рассматривается простая замена k исходных факторов на k новых (аффинное преобразование типа поворота осей координат), то в факторном анализе количество новых факторов f_i ($i = 1, 2, \dots, m$) меньше количества исходных: $m < k$. Тогда исходные наблюдаемые факторы можно представить через новые в виде:

$$x_j = \sum_{i=1}^m l_{ji} f_i + \varepsilon_j, \quad j = 1, 2, \dots, k; \quad i = 1, 2, \dots, m,$$

где ε_j – остаточная случайная поправка, действующая только на исходный, наблюдаемый фактор x_j , но не на новые f_i . Такое представление следует понимать, как предположение о наличии ненаблюдаемых, некоррелированных, "высших" факторов f_i , которые "объясняют" все наблюдаемые x_j (от которых зависят все наблюдаемые). Т.е. можно сформулировать и решить задачу минимизации количества независимых (некоррелированных) факторов f_i .

На практике, однако, не решают такую сложную задачу оптимизации, а просто подбирают наиболее приемлемый набор "высших" факторов. Конечно, выбор здесь неоднозначен, поэтому пытаются отыскать наиболее легко интерпретируемые, комплексные факторы. Собственно *факторным анализом* и является такой поиск.

4) Существует еще один способ выявления наиболее значимых факторов изучаемого процесса – планирование отсеивающих экспериментов, состоящее в постепенном отсеивании несущественных факторов из первоначально выбранного большого их числа. Такое планирование можно осуществлять даже в условиях сверхнасыщенных планов, в которых число опытов меньше числа изучаемых факторов и, соответственно, числа коэффициентов регрессионной модели. Наиболее распространенным методом для этого является метод случайного баланса, в котором матрица плана с числом опытов N , меньшим числа факторов k , строится из случайных наборов уровней по столбцам. Можно строить матрицу и из случайного набора дробных реплик. При этом предполагается, что степень влияния факторов, расположенных в порядке **убывания**, представляется убывающей экспонентой. Задача отсеивающего эксперимента состоит в том, чтобы провести ранжирование всех факторов.

5) Из нескольких методических походов к планированию эксперимента для "хорошо организованных систем" рассмотрим задачу **уточнения** параметров модели, которую в теории математического моделирования называют задачей *идентификации* (§ 2.3). Примером такой задачи может служить уточнение характеристик летательного аппарата по результатам летных испытаний или расследование причин летных происшествий. В этих случаях считается априори известным аналитический вид функциональной зависимости выходного параметра от входных:

$$y = \varphi(x_1, x_2, \dots, x_k; \lambda_1, \lambda_2, \dots, \lambda_s) = \varphi(\mathbf{X}; \mathbf{\Lambda}),$$

где $\mathbf{\Lambda} = (\lambda_1, \lambda_2, \dots, \lambda_s)$ – вектор параметров модели, для которых известны предварительные оценки их значений: $\mathbf{\Lambda}^{(0)} = (\lambda_1^{(0)}, \lambda_2^{(0)}, \dots, \lambda_s^{(0)})$. Задача формулируется так: необходимо подобрать план эксперимента из N опытов $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{ik})$, где $i = 1, 2, \dots, N$ чтобы по их результатам можно было бы уточнить параметры модели наилучшим образом. Под наилучшим здесь обычно понимают минимизацию обобщенной дисперсии оценок, т.е. критерий D-оптимальности (см. § 7.1).

8.2. Методы экспертных оценок

Методы экспертных оценок занимают в планировании эксперимента особое место благодаря тому, что их можно применять на любой стадии исследовательской работы и в самых разнообразных целях. Здесь следует упомянуть лишь самые распространенные вопросы, которые решаются с их помощью:

- формулировка постановки задачи исследований;
- выбор основных факторов;
- выбор наблюдаемых и регистрируемых параметров;
- структурирование сложных систем;
- ранжирование критериев и целей деятельности отдельных элементов сложных систем и т.д.

Группы экспертов (от 1 до 30 специалистов) создаются с учетом их квалификации по специально разработанным методикам.

Основным требованием всех методов экспертных оценок является строгая формулировка вопросов экспертизы, содержащих единообразно понимаемые и единичные (не распространенные или комплексные) темы.

Ниже, без дополнительных рассуждений, приводится краткий обзор основных аспектов теории экспертных оценок.

Задачи экспертизы:

1. Оценивание имеющихся объектов по качественным и количественным характеристикам.
2. Построение множества объектов (составление списка).
3. Построение множества объектов и их оценивание.

Этапы экспертизы:

- формирование множества оцениваемых объектов (вопросов);
- формирование экспертной комиссии;
- выбор процедуры экспертного оценивания (выбор метода);
- обработка результатов экспертного оценивания;
- анализ результатов экспертизы и принятие решения.

Классификация методов экспертизы:

- по количеству экспертов:
 - индивидуальные,
 - коллективные (следует учитывать, что при количестве экспертов менее 7 информация страдает неполнотой, а стабилизируются результаты экспертизы при 10 – 30 экспертах);
- по методу опроса (возможны сочетания методов):
 - индивидуальный,
 - очный (интервью) с обратной связью или без нее,
 - открытый (ответ в произвольной форме),
 - закрытый (ответ типа да/нет или выбор ответа из предложенного набора),
 - прямой (до экспертов доводится цель экспертизы),
 - косвенный (до экспертов не доводится цель экспертизы);

- по информированности экспертов:
 - с заданным множеством объектов оценивания,
 - с неопределенным множеством объектов оценивания;
- по организации процедуры:
 - простые,
 - сложные (с декомпозицией проблемы на простые вопросы);
- по типу оценок:
 - качественные,
 - количественные.

Шкалы экспертных оценок:

- абсолютная шкала (измерение в принятых единицах) – позволяет получить самую тонкую количественную оценку;
 - шкала интервалов (в принятых единицах выбор интервала, в котором находится оцениваемый параметр) – дает "мягкую" количественную оценку;
 - шкала отношений (измерение "во сколько раз") – дает относительную количественную оценку;
 - шкала разностей (измерение в принятых единицах "на сколько" больше или меньше) – дает разностную количественную оценку;
 - шкала порядка (распределение по порядку, рангу без учета кратности предпочтения) – дает качественную оценку;
 - номинальная шкала (разбиение на классы эквивалентности) – дает качественную оценку;
 - балльная шкала (оценка в баллах по принятой размерности: непрерывно – от 0 до 1, или дискретно – целые от 0 до 5, или 10, или 100).

Процедуры экспертного оценивания альтернатив:

- ранжирование (наиболее предпочтительная оценка имеет ранг $R = 1$, остальные – бóльшие целые значения – ранг определяется последовательной нумерацией альтернатив в порядке убывания предпочтения);
- парное сравнение (построение матрицы парных сравнений элементов:

$$a_{ij} = \begin{cases} 0 & \text{– при } \alpha_i < \alpha_j, \\ 1 & \text{– при } \alpha_i \approx \alpha_j, \\ 2 & \text{– при } \alpha_i > \alpha_j, \end{cases}$$

где α_i – значение сравниваемого признака i -го элемента);

- метод фон Неймана-Моргенштерна (матрица парных сравнений имеет непрерывные значения элементов $0 \leq \alpha_i \leq 1$);
- непосредственная оценка (дается балльная оценка (см. выше) каждой альтернативе, для эквивалентности используется одно общее значение);
- последовательное сравнение – метод Чёрчмена-Акофа – по рекуррентному алгоритму:

- 1 ► ранжирование альтернатив,
- 2 ► присвоение "1" первой альтернативе, остальным меньше 1,

3► сравнение предпочтительности первой альтернативы перед **всеми следующими вместе** и при необходимости увеличение (или уменьшение) оценки "1" до величины, большей (меньшей) суммы остальных,

4► повторение сравнения для второй и последующих альтернатив.

Методы проведения экспертизы (формы организации экспертизы)

Метод мозговой атаки (метод мозгового штурма) – коллективный открытый метод, при котором на поставленную цель экспертизы эксперты отвечают генерацией произвольных идей: без критики, без аргументации, может быть с развитием или с фантазией. Далее идеи лаконизируются; не соответствующие цели, отсеиваются; систематизируются; рассматриваются прямо противоположные; оценивается возможность реализации. В заключение, после отбора руководителями экспертизы нескольких идей, они могут подвергаться всей группой экспертов конструктивной критике по **конкретным** позициям: логичность, этичность, эстетичность, экологичность, социальность и т.п.

Метод дискуссии – коллективный открытый метод, при котором во время демократического обсуждения заданной тематики происходит генерация идей, их формулировка и обобщение.

Метод комиссий – коллективный открытый метод, при котором предусматривается предварительный отбор состава экспертов и разбиение их на комиссии по направлениям, предоставление всей требуемой информации и в заключение коллегиальная оценка по принятой шкале.

Метод суда – метод, близкий к методу комиссий, но состав отбирается не административным образом, а с обеспечением противоположных мнений и интересов, и информация предполагается субъективной.

Метод деловых совещаний – коллективный открытый метод, при котором ведущий формулирует проблему, участники высказывают дополнительную информацию, альтернативы, идеи и оценки, а окончательный выбор решения или оценки делает и оглашает ведущий.

Анкетирование и интервьюирование – метод, который может быть любым по количеству экспертов и по методу опроса, предъявляет особые требования к лаконичности и однозначности вопросов и ответов, проводится в письменной или устной форме чаще всего независимым опросом экспертов.

Метод Дельфи – анонимное анкетирование в несколько этапов (2 – 3) с обратной связью, когда всем экспертам сообщаются результаты обработки (но не частные мнения) ответов предыдущего этапа с аргументацией крайних суждений.

Метод решающих матриц – предполагает декомпозицию сложной проблемы на несколько уровней, в которых возможны альтернативные решения, связанные с решениями на соседних уровнях; эксперты строят в конечном итоге матрицу оценок, сумма которых равна 1.

Метод прогнозного графа – многошаговый метод, предполагающий: составление списка целей, систематизацию их подчиненности, оценку возможности реализации подцелей и их уточнение (можно проводить другим составом

экспертов), спуск по дереву целей до таких "заземленных", решение которых очевидно.

Очевидно, что перечисленные методы могут комбинироваться друг с другом.

Обработка результатов экспертизы может быть проведена по следующей схеме, отдельные элементы которой могут не использоваться.

1) Показатели обобщенного мнения

1.1) Средняя балльная оценка $M_j = \frac{1}{m_j} \sum_{i=1}^{m_j} C_{ij}$, где m_j – число экспертов, оценивших j -й вопрос, C_{ij} – оценка по балльной шкале $0 \leq C \leq B$ (где B – высшая балльная оценка в принятой шкале), данная i -ым экспертом j -ому вопросу ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$).

1.2) Частота k -ой оценки j -го вопроса $K_j^k = \frac{m_j^k}{m_j}$.

1.3) Сумма рангов j -го вопроса $S_j = \sum_{i=1}^{m_j} R_{ij}$, где R_{ij} – ранг.

2) Показатели степени согласованности экспертов

2.1) Коэффициент вариации мнений по j -му вопросу $V_j = \frac{\sigma_j}{M_j}$, где

$\sigma_j = \sqrt{\frac{1}{m_j-1} \sum_{i=1}^{m_j} (C_{ij} - M_j)^2}$; согласованность тем выше, чем меньше V_j .

2.2) Коэффициент конкордации $W = \frac{12}{m^2(n^3 - n) - m \sum_{i=1}^m T_i} \cdot \sum_{j=1}^n d_j^2$, где

$d_j = S_j - \frac{1}{n} \cdot \sum_{j=1}^n S_j$; $T_i = \sum_{l=1}^L (t_l^2 - t_l)$, t_l – кратность l -го ранга; полная согласованность экспертов – при $W = 1$, согласованности нет при $W = 0$.

3) **Статистическая значимость показателя согласованности W** по критерию Пирсона: $\chi_n^2 = \frac{1}{mn(n+1) - \frac{1}{n} \sum_{i=1}^m T_i} \cdot \sum_{j=1}^n d_j^2$ с $n-1$ степенями свободы.

4) **Коэффициент активности эксперта** при ответе на j -ый вопрос:

$$K_{\text{э}j} = \frac{m_j}{m}$$

5) **Учет компетентности экспертов** может быть проведен с помощью следующих оценок.

5.1) Самооценка по балльной или шкальной системе таких объективных показателей, как уровень, широта и стаж специализации по вопросам экспертизы и смежным. Самооценка производится по специальным таблицам, не допус-

кающим произвола и задающим каждому уровню фактора свою балльную оценку, например, по балльной шкале табл. 39 с вычислением коэффициента осведомленности $K_{oc} = B / 10$.

Таблица 39.

| Градация | Балл |
|--|---------|
| не знаком с вопросом | 0 |
| плохо знаком, но вопрос входит в сферу знаний и деятельности | 1, 2, 3 |
| удовлетворительно знаком, тесно связан со смежной узкой специализацией | 4, 5, 6 |
| хорошо знаком, принимает практическое участие | 7, 8, 9 |
| вопрос входит в узкую специализацию | 10 |

5.2) Самообъективная оценка вычисляется, как сумма оценок степени влияния всех видов источников информации по шкале табл. 40 (максимальное значение равно 1).

Таблица 40.

| Источники информации | Степень влияния источника | | |
|---|---------------------------|---------|--------|
| | высокая | средняя | низкая |
| производственный опыт | 0,5 | 0,4 | 0,2 |
| проведенный теоретический анализ | 0,3 | 0,2 | 0,1 |
| обобщение отечественной литературы | 0,05 | 0,04 | 0,03 |
| обобщение зарубежной литературы | 0,05 | 0,04 | 0,03 |
| личное знакомство с состоянием вопроса за рубежом | 0,05 | 0,04 | 0,03 |
| интуиция | 0,05 | 0,04 | 0,03 |

5.3) Коэффициент компетентности i -го эксперта ($i = 1, 2, \dots, m$) по j -ому вопросу ($j = 1, 2, \dots, n$) может быть вычислен по формуле:

$$k_{ij} = \frac{\sum_{v=1}^s \theta_{vi} R_{vj}}{\sum_{v=1}^s R_{vj}},$$

составляющие которой определяются следующей процедурой. Единым образом (руководителем, или каждым экспертом для себя, или взаимно) для каждого i -го эксперта определяются уровни осведомленности $0 < \theta_{vi} < 10$ (целые числа) по каждой v -ой из s общепризнанных специализаций (дисциплин, сфер деятельности), к которым можно отнести все вопросы экспертизы. Каждый j -й во-

прос экспертизы получает оценку R_{vj} ($0 \leq R_{vj} \leq 2$) по уровню важности для него каждой v -ой специализации с помощью двухбалльной непрерывной шкалы табл. 41.

Таблица 41.

| Уровень важности специализации для вопроса | Оценка |
|--|--------|
| весьма существенно | 2 |
| полезно | 1 |
| несущественно | 0 |

5.4) Исходная оценка (как в п. 5.6 и 5.7) на основе информации о **предыдущих** экспертизах и **знании истины**.

5.5) Взаимная оценка экспертов (исключая самооценку).

5.6) Мгновенная оценка по итогам данной экспертизы: $K_{ij} = 1 - \frac{|C_{ij} - M_j|}{B}$

– по j -ому вопросу i -го эксперта, $K_i = \frac{1}{n} \cdot \sum_{j=1}^n K_{ij}$ – по всем вопросам i -го эксперта.

5.7) Итоговая оценка на основании знания об истине: $\bar{K}_{ij} = 1 - \frac{|C_{ij} - \bar{M}_j|}{B}$,

где \bar{M}_j – истинное значение балльной оценки.

Список литературы

1. Альсведе Р., Вегенер И. Задачи поиска. – М.: "Мир", 1982. – 368 с.
2. Барзилович Е.Ю. Оптимально управляемые случайные процессы и их приложения (теоретические основы эксплуатации авиационных систем по состоянию). – Егорьевск: ЕАТК ГА, 1996. – 299 с.
3. Белов В.В., Воробьев Е.М., Шаталов В.Е. Теория графов. Учебное пособие для вузов. – М.: Высшая школа, 1976. – 392 с.
4. Березин И.С., Жидков Н.П. Методы вычислений. Том 1. – М.: Наука, 1966. – 632 с.
5. Бернацкий Ф.И. Планирование экспериментов в инженерных исследованиях. – Владивосток: 1986. – 45 с.
6. Бормотов М.Ю., Гуров А.Г., Корунов С.С., Кукушкин С.Н. Экспертные методы прогнозирования. – М.: МАИ, 1985. – 60 с.
7. Васильев Ф.П. Численные методы решения экстремальных задач. – М.: Наука, 1980. – 520 с.
8. Вентцель Е.С. Исследование операций: задачи, принципы, методология. – М.: Наука, 1980. – 208 с.
9. Вентцель Е.С. Теория вероятностей. – М.: Наука, 1964. – 576 с.
10. Вилисов В.Я. и др. Экспертные методы в АСУ производством и обработкой ЛА. – М.: МАИ, 1984. – 72 с.
11. Годунов С.К., Рябенский В.С. Разностные схемы (введение в теорию). – М.: Наука, 1973. – 400 с.
12. ГОСТ 24026–80. Исследовательские испытания. Планирование эксперимента. Термины и определения. – М.: Изд-во стандартов, 1980.
13. Добров Г.М., Ершов Ю.В., Левин Е.И., Смирнов Л.П. Экспертные оценки в научно-техническом прогнозировании. – Киев: Наукова Думка, 1974. – 160 с.
14. Дыхненко Л.М. и др. Основы моделирования сложных систем: Учебное пособие для вузов. – Киев: Вища школа. 1981. – 359 с.
15. Ибрагимов И.А. и др. Моделирование систем: Учебное пособие. – Баку: Азинефтехим, 1989. – 83 с.
16. Корн Г., Корн Т. Справочник по математике (для научных работников и инженеров). – М.: Наука, 1973. – 832 с.

17. Красовский Г.И., Филаретов Г.Ф. Планирование эксперимента. – Минск: БГУ, 1982. – 302 с.
18. Кубланов М.С. Планирование экспериментов и обработка результатов: Учебно-методическое пособие по изучению дисциплины и варианты заданий РГР. – М.: МГТУ ГА, 1998. – 36 с.
19. Лебедев А.Н. Моделирование в научно-технических исследованиях. М.: Радио и связь, 1989. – 224 с.
20. Липатов Е.П. Теория графов и ее применения. – М.: Знание, 1986. – 32 с.
21. Мышкис А.Д. Элементы теории математических моделей. – М.: Физматгиз, 1994. – 192 с.
22. Налимов В.В. Теория эксперимента. – М.: Наука, 1971. – 208 с.
23. Неймарк Ю.И., Коган Н.Я., Савелов В.П. Динамические модели теории управления. – М.: Наука, 1995. – 400 с.
24. Остославский И.В., Стражева И.В. Динамика полета. Траектории летательных аппаратов. – М.: Машиностроение, 1969. – 500 с.
25. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений. – М.: Наука, 1968. – 288 с.
26. Савченко А.А. Введение в математическую статистику с применением в гражданской авиации. – Киев: МИИГА, 1975. – 132 с.
27. Савченко А.А. Многомерный статистический анализ для инженеров гражданской авиации. – М.: МИИГА, 1976. – 112 с.
28. Советов Б.Я., Яковлев С.Я. Моделирование систем: Учебник для вузов. – М.: "Высшая школа", 1998. – 320 с.
29. Хальд А. Математическая статистика с техническими приложениями. – М.: Изд-во иностранной литературы, 1956. – 664 с.
30. Хикс Ч.Р. Основные принципы планирования эксперимента. – М.: Мир, 1967. – 406 с.
31. Чисар И., Кёрнер Я. Теория информации: теоремы кодирования для дискретных систем без памяти. – М.: Мир, 1985. – 400 с.
32. Шилейко А.В., Кочнев В.Ф., Химушин Ф.Ф. Введение в информационную теорию систем. – М.: Радио и связь, 1985. – 280 с.
33. Шторм Р. Теория вероятностей. Математическая статистика. Статистический контроль качества. – М.: Мир, 1970. – 368 с.