ВСЕСОЮЗНАЯ АКАДЕМИЯ с.-х. НАУК им. В. И. ЛЕНИНА
ВСЕСОЮЗНЫЙ ИНСТИТУТ ЗАЩИТЫ РАСТЕНИЙ

ПРОФ. Ю. Л. ПОМОРСКИЙ

СТАТИСТИЧЕСКИЙ АНАЛИЗ КОМПЛЕКСА ПРИЗНАКОВ

(Конспект лекций)

ИЗДАНИЕ ВСЕСОЮЗНОГО ИНСТИТУТА ЗАЩИТЫ РАСТЕНИЙ
ЛЕНИНГРАД 1938

Отв. ред. А. Н. Касихин — Тех. ред. А. А. Дмитриев
Сдано в набор 30/I 1938 г. — Подписано к печати 13/IV 1938 г.
Ленгорлит № 749 — Заказ № 241 — Формат 62 × 94
Объем 6 авт. л., 2¾ бум. л. — 105.000 тип. зн. в бум. л. — Тираж 1000 экз.

Ленпромпечатьсоюз, типография арт. „Печатня", Ленинград, Прачечный пер., 6

От автора

Настоящая работа представляет собою конспект краткого цикла лекций, прочитанных автором в 1937—38 уч. году научным сотрудникам и аспирантам Всесоюзного Института Защиты Растений, издавшего этот конспект, и слушателям Высших Агро-Метеорологических Курсов при Всесоюзном Институте Агро-Гидро-Метеорологии, по предложению которого настоящий конспект был составлен. Кроме того, данная работа содержит материалы, частично изложенные автором в лекциях, прочитанных им в разное время студентам, аспирантам и научным сотрудникам Ленинградского Сельскохозяйственного Института, Физиологического Института Академии Наук, Биологической станции ВИЭМ, Института Мозга, Института Охраны Здоровья Детей и Подростков, Акушерско-Гинекологического Института и на Высших Курсах Прикладной Зоологии и Фитопатологии.

В основу этих лекций были положены две известные работы R. A. Fisher'a: 1) „Statistical Methods for Research Workers" 1936 и 2) The Design of Experiments" 1935, а также некоторые более ранние и позднейшие работы Fisher'a, „Student'a" и др. авторов по теории и методике так называемой „малой выборки" и в том числе дополнительные главы к русскому переводу книги W. Johannsen'a „Элементы точного учения об изменчивости и наследственности", написанные Н. Ф. Деревицким. С отдельными приемами новейшей методики Fisher'a автор познакомился на лекциях д-ра C. I. Bliss'a, в которых он излагал материалы, заимствованные непосредственно из лекций R. A. Fisher'a. Некоторые методы и схемы разработаны автором настоящего конспекта и публикуются впервые.

Следует отметить, что этот конспект первоначально не предназначался для печати и был составлен в кратком и не вполне систематизированном виде для размножения на ротаторе в ограниченном числе экземпляров. Отсюда понятна некоторая неровность и неполнота изложения и ряд чисто редакционных недочетов, за которые автор приносит свои извинения.

В настоящую работу, в силу ограниченности ее объема и связанности програмой курсов АГМИ, вовсе не вошли некоторые материалы (напр. описание „смешанной схемы“, так наз. „Analysis of Covariance“ и др.). Автор надеется в ближайшее время дать эти дополнения в форме продолжения настоящего конспекта (его II части).

Пользующимся настоящим конспектом рекомендуется при первом чтении пропустить главы X—XVIII (стр. 38—64).

Проф. *Ю. Л. Поморский*

Статистический анализ комплекса признаков

I

Задача, указанная в заголовке, имеет чрезвычайно широкое применение в самых разнообразных отраслях использования статистического метода. Агроном-опытник стремится выяснить влияние ряда метеорологических факторов или агротехнических приемов на урожай определенной культуры; зоотехник исследует влияние различных кормов и рационов на увеличение веса животного; фармаколог изучает действие различных лекарственных веществ и их дозировок на функцию того или иного органа; клиницист накапливает наблюдения над действием различных лечебных средств и особенностей режима на протекание той или иной болезни; физиолог изучает изолированное и совокупное действие ряда раздражителей на силу или продолжительность ответной реакции; психолог-экспериментатор исследует влияние цвета, размера и формы предмета на те или иные особенности его восприятия; педагог изучает эффективность отдельных методов обучения в соответствии с индивидуальными особенностями учащихся и специфическим характером предмета преподавания; методист-физкультурник ставит опыты над влиянием того или иного комплекса физических упражнений на здоровье и работоспособность занимающихся; профгигиенист исследует всю совокупность условий, обеспечивающих наилучшую производительность труда при минимальной его вредности и т. д. Во всех этих примерах мы имеем однотипную схему: требуется выяснить, как влияет совокупность факторов (A, B, C, D и т. д.) на тот или иной интересующий исследователя количественный результативный признак (V).

При анализе этой схемы необходимо различать два случая: 1) дозировка всех факторов (A, B, C и т. д.) полностью зависит от экспериментатора (например, при испытании сроков посева, рационов корма, силы раздражителей, дозировки лекарственных средств или продолжительности физических упражнений), в силу чего он получает возможность произвольно комбинировать эти факторы, изучая действие всех без исключения их сочетаний, и 2) наблюдатель самостоятельно условий опыта не создает и лишь пассивно регистрирует те конкретные сочетания изучаемых факторов и их дозировок, которые сами собой обнаружатся в его опыте (напр., те или иные сочетания метеорологических факторов, симптомов болезни, индивидуальных особенностей испытуемых и т. д.). Первый случай мы будем именовать „схемой эксперимента“, а второй — „схемой наблюдения“.

Ниже дается числовой пример математико-статистического анализа изолированного и совокупного действия трех факторов: A, B и C на количественный результативный признак V по первой „схеме эксперимента“.

Пусть задание сведено в следующую таблицу:

		C_1	C_2	C_3	C_4
A_1	B_1	7	9	8	12
	B_2	4	12	7	5
	B_3	10	9	3	10
A_2	B_1	11	19	14	12
	B_2	14	13	7	14
	B_3	11	13	12	16
A_3	B_1	12	11	5	12
	B_2	6	8	7	11
	B_3	6	14	9	7

Таблицу эту можно для наглядности иллюстрировать различными наименованиями в зависимости от области ее применения. Так, в частности, она может изображать результат испытания урожайности трех каких-либо культур: A_1, A_2 и A_3, при трех сроках посева: B_1, B_2 и B_3, на четырех участках поля: C_1, C_2, C_3 и C_4. В этом случае числа, записанные внутри таблицы (так наз. „даты“ V) будут характеризовать урожай, собранный с отдельных делянок. В применении к физиологическому эксперименту эта схематическая таблица могла бы изображать результаты испытания трех каких-либо химических агентов A в трех дозировках B при введении их в кровь четырех опытных животных, а даты V характеризовать отдельные результаты этих опытов (напр., продолжительность процессов, вызываемых действием этих агентов).

Обозначим число вариантов факторов A, B и C соответственно буквами a, b и c (у нас, следовательно, $a = 3$, $b = 3$, а $c = 4$), а общее число всех дат V (делянок, отдельных опытов) обозначим буквою N (у нас $N = 36$).

Для возможности применения описанных ниже методов необходимо строгое соблюдение следующих двух условий: 1) в данной серии экспериментов должны быть испытаны все без исключения сочетания каждого варианта отдельного фактора со всеми вариантами остальных факторов, в силу чего общее число всех дат N окажется равным произведению чисел: a, b и c ($N = abc = 3.3.4 = 36$), 2) отдельные даты V могут представлять собою как результаты единичных измерений, так и средние из нескольких измерений, но в этом последнем случае все они должны быть получены из одинакового числа отдельных наблюдений.

Математико-статистическая обработка этих данных протекает в двух последовательных стадиях: 1) анализ действия отдельных факторов A, B и C в целом (изолированно, совместно и по сравнению друг с другом) и 2) детальный анализ действия отдельных вариантов каждого из этих факторов.

Первая фаза этой разработки начинается с подсчета средних результатов испытания по каждому отдельному варианту всех факторов и общего среднего (M) для всего опыта в целом. Для этого удобно просуммировать предварительно все $N=36$ выписанных в таблице дат V (как по отдельным строчкам, так и по всем столбцам):

		C_1	C_2	C_3	C_4	
A_1	B_1	7	9	8	12	**36**
	B_2	4	12	7	5	**28**
	B_3	10	9	3	10	**32**
A_2	B_1	11	19	14	12	**56**
	B_2	14	13	7	14	**48**
	B_3	11	13	12	16	**52**

следующие частные средние (суммируются ранее уже подсчитанные итоги соответствующих строк):

$$B_1 = \frac{36+56+40}{12} = \frac{132}{12} = 11$$

$$B_2 = \frac{28+48+32}{12} = \frac{108}{12} = 9$$

$$B_3 = \frac{32+52+36}{12} = \frac{120}{12} = 10.$$

Проверка этих средних дает следующий результат:

$$\frac{11+9+10}{3} = \frac{30}{3} = 10 = M.$$

В применении к четырем вариантам последнего фактора C (средняя урожайность на отдельных участках поля, зависящая от особенностей почвы, или характеристики индивидуальных свойств отдельных опытных животных) получим:

$$C_1 = \frac{81}{9} = 9 \qquad C_3 = \frac{72}{9} = 8$$

$$C_2 = \frac{108}{9} = 12 \qquad C_4 = \frac{99}{9} = 11$$

Проверка дает:

$$\frac{9+12+8+11}{4} = \frac{40}{4} = 10 = M.$$

II

Проглядывая все $N = 36$ первоначальных дат V и все только что вычисленные по ним частные средние, мы обнаруживаем заметную их колеблемость (варьирование, изменчивость, рассеяние) около общего среднего $M = 10$. Измерим величину этой колеблемости и постараемся выяснить ее механизм.

Возьмем какую-нибудь отдельную дату V, например $V = 19$ (4-я строчка сверху и третий столбец справа). Этот частный результат одного из 36 измерений отклоняется от общего среднего $M = 10$ на $+9$ единиц. Обозначив это отклонение буквой x, будем иметь:

$$x = V - M = 19 - 10 = +9$$

Для другой какой-либо даты, напр. $V = 3$ (3-я строчка сверху и 2-й столбец справа), ее отклонение:

$$x = V - M = 3 - 10 = -7.$$

Вычислив такие отклонения (x) для всех 36 дат V, мы всегда получим в сумме нуль ($\Sigma x = 0$). Отсюда заключаем, что сумма этих отклонений не годится для характеристики „общего рассеяния“ всех дат V около их среднего M. Для этой цели применяется сумма квадратов отдельных отклонений (Σx^2). В нашем примере для всех 36-ти дат V сумма квадратов их отклонений от общего среднего $M = 10$:

$$\Sigma x^2 = 450,$$

в чем легко убедиться, проделав соответствующие вычисления в действительности.

Эта сумма (450) и может служить искомой числовой характеристикой общего рассеяния всех первоначальных дат V.

Аналогично найдем сумму квадратов отклонений каждой отдельной серии частных средних около того же общего среднего M. Выше мы получили три частные средние по отдельным вариантам фактора A ($A_1 = 8$, $A_2 = 13$ и $A_3 = 9$). Отклонения этих частных средних от общего среднего $M = 10$ мы обозначим греческими буквами α (альфа). Вот расчет этих отклонений и их квадратов:

A	α	α^2
8	-2	4
13	$+3$	9
9	-1	1
Сумма	0	14

Здесь сумма этих отклонений:
$$\Sigma \alpha = 0,$$
а сумма их квадратов:
$$\Sigma \alpha^2 = 14.$$

В применении ко второму фактору B получим следующий расчет отклонений β (бэта) и их квадратов:

B	β	β^2
11	$+1$	1
9	-1	1
10	0	0
Сумма	0	2

$$\Sigma \beta = 0$$

$$\Sigma \beta^2 = 2.$$

Наконец, для третьего фактора C получим следующие значения отклонений γ (гамма) и их квадраты:

C	γ	γ^2
9	-1	1
12	$+2$	4
8	-2	4
11	$+1$	1
Сумма	0	10

$$\Sigma \gamma = 0$$

$$\Sigma \gamma^2 = 10.$$

Эти три суммы: $\Sigma \alpha^2 = 14$, $\Sigma \beta^2 = 2$ и $\Sigma \gamma^2 = 10$ могут служить числовой характеристикой колеблемости соответствующих серий частных средних (A, B и C) около общего среднего M. Однако в таком своем виде эти суммы не являются все же сравнимыми ни между собой, ни с ранее вычисленной суммой $\Sigma x^2 = 450$, поскольку первые две суммы ($\Sigma \alpha^2$ и $\Sigma \beta^2$) были получены нами всего лишь для $a = 3$ и $b = 3$ частных средних каждая, а третья сумма ($\Sigma \gamma^2$) для $c = 4$ частных средних, в то время как сумма Σx^2 была вычислена выше для всех $N = 36$ первоначальных дат V. Для приведения этих сумм к сравнимому виду будем рассуждать так: комплект из 3-х частных средних (A_1, A_2 и A_3) обнаруживает колебания, измеряемые суммою $\Sigma \alpha^2 = 14$, в то время как наши 36 дат дают колебания, измеряемые суммой $\Sigma x^2 = 450$. Всех дат здесь в 12 раз больше, чем этих частных средних, поэтому и сумму $\Sigma \alpha^2 = 14$ для сопоставления ее с суммой $\Sigma x^2 = 450$ следует предварительно увеличить в 12 раз. Помножив 14 на 12, получим 168. Аналогично найдем, что если бы частных средних по отдельным вариантам фактора B (B_1, B_2 и B_3) было не 3 ($b=3$), а столько же, сколько у нас было и первоначальных дат V ($N=36$), т. е. также в 12 раз больше, то при той же их колеблемости мы получили бы сумму квадратов отклонений ($\Sigma\beta^2$) не $=2$, а в 12 раз больше, т. е. $=24$. Наконец, в применении к фактору C, имеющему $c=4$ варианта, соответствующую сумму квадратов отклонений $\Sigma\gamma^2$ следует увеличить в 9 раз, в силу чего эта сумма (10) окажется равной 90. Вот эти три новые суммы (168, 24 и 90), приведенные к одному и тому же числу слагаемых (всюду к 36), будут уже сравнимы как между собой, так и с суммой $\Sigma x^2=450$.

В дальнейшем все эти приведенные суммы мы будем обозначать буквами S с соответствующими значками: x, a, b и c и, чтобы не повторять каждый раз всех этих рассуждений, вычислять их по следующим окончательным формулам:

$$S_a=\frac{N\Sigma\alpha^2}{a}=\frac{36\cdot 14}{3}=168$$

$$S_b=\frac{N\Sigma\beta^2}{b}=\frac{36\cdot 2}{3}=24$$

$$S_c=\frac{N\Sigma\gamma^2}{c}=\frac{36\cdot 10}{4}=90$$

$$S_x=\Sigma x^2=450.$$

Теперь займемся анализом причин, вызывающих колебания отдельных дат V, для чего рассмотрим какую-либо определенную дату, напр. $V=19$ (4-я строчка сверху и 3-й столбец справа). Эта дата превышает общее среднее $M=10$ на значительную величину:

$$x=V-M=19-10=+9.$$

Спрашивается: почему она получила здесь такое сильное положительное отклонение $x=+9$? Ответ ясен. Эта дата ($V=19$) возникла здесь под влиянием чрезвычайно благоприятного сочетания отдельных вариантов всех трех факторов A, B и C. В применении к схеме полевого опыта эта дата характеризовала бы сравнительно очень высокий урожай, собранный с делянки, засеянной сортом A_2, в срок B_1, на участке C_2 (см. таблицу). Но мы уже знаем, что этот сорт при прочих равных условиях давал самый высокий средний урожай $A_2=13$; данный срок посева, по сравнению с другими, также характеризовался самым большим значением частного среднего $B_1=11$, и соответствующий участок поля по своим почвенным условиям оказался самым благоприятным, давшим на всех своих делянках средний урожай $C_2=12$. В применении к физиологической интерпретации этой схемы мы могли бы объяснить столь высокое положительное отклонение даты $V=19$ тем, что здесь действовал чрезвычайно активный химический агент A_2 в наиболее эффективной дозировке B_1 на сильно реактивный объект C_2.

Для этих трех частных средних $A_2 = 13$, $B_1 = 11$ и $C_2 = 12$ их отклонения (α, β и γ) от общего среднего $M = 10$ соответственно равны:

$$\alpha = +3 \qquad \beta = +1 \qquad \gamma = +2.$$

Следовательно, данный вариант фактора A, вообще говоря, повышал величину отдельных дат в среднем на $\alpha = +3$, данный вариант фактора B сам по себе мог бы (в среднем) повысить ту или иную дату V на величину $\beta = +1$, а рассматриваемый вариант фактора C способен увеличить ее на $\gamma = +2$. Вполне естественно, что все эти факторы, действуя совместно, дали некоторый суммарный эффект, выразившийся здесь в заметном увеличении этой даты $V = 19$ по сравнению с общим средним $M = 10$. Однако действительная сумма этих трех отклонений:

$$\alpha + \beta + \gamma = +3 + 1 + 2 = +6$$

все же не равна здесь фактическому отклонению этой даты $V = 19$ от общего среднего $M = 10$:

$$x = V - M = 19 - 10 = +9.$$

Чем же объясняется эта разница ($9 - 6 = +3$)? Очевидно тем, что на результативный признак V, кроме трех учтенных в этом опыте факторов A, B и C, влияют еще множество других неучтенных факторов, как независимых, так и связанных между собой, и в различных беспорядочных своих сочетаниях более или менее сильно изменяющих все наши даты V. Эти хаотические до конца неустранимые влияния мы условно назовем „случайными", а зависящие от них дополнительные отклонения отдельных дат V обозначим буквой z.

В нашем примере (для $V = 19$) все три учтенные фактора дают сумму отклонений $\alpha + \beta + \gamma = +6$, а недостающее до $x = +9$ отклонение $9 - 6 = +3$ и объясняется действием всей совокупности таких „случайных" влияний. Следовательно, здесь „случайное" отклонение $z = +3$.

Итак имеем равенство:

$$x = \alpha + \beta + \gamma + z.$$

Если бы мы могли полностью изолировать результативный признак V от каких бы то ни было воздействий на него со стороны всех „случайных" факторов, то величина отдельных дат V (а стало быть и их отклонений x от общего среднего M) зависела бы только от трех изучаемых факторов: A, B и C, а все случайные отклонения z оказались бы равными 0. Так, в частности для $V = 19$ при этих условиях мы имели бы:

$$x = \alpha + \beta + \gamma = +3 + 1 + 2 = +6,$$

т. е. эта дата отклонялась бы от среднего $M = 10$ не на $+9$, а на $+6$, т. е. оказалась бы равной не 19, а только 16. Эту „теоретическую" дату (16) в отличие от „эмпирической" (19) мы обозначим буквой W; итак у нас:

$$V = 19, \text{ а } W = 16.$$

В отдельных случаях теоретические даты W представляется удобным вычислять по формуле:

$$W = A + B + C - (f - 1)\ M,$$

где f — общее число изучаемых факторов (у нас $f = 3$). Следовательно, в применении к нашей схеме:

$$W = A + B + C - 2\ M.$$

Так как эмпирическая дата $V = 19$ возникла под действием сочетания факторов $A_2 = 13$, $B_1 = 11$ и $C_2 = 12$ (см. таблицу), то соответствующая ей теоретическая дата:

$$W = 13 + 11 + 12 - 2 . 10 = 16,$$

что мы имели и раньше.

Очевидно, случайное отклонение z является разностью между эмпирической и теоретической датой:

$$z = V - W.$$

Пользуясь формулой для вычисления W, мы могли бы для всех 36 эмпирических дат V получить соответствующие им теоретические значения. Выполнив эти вычисления в действительности и записав полученные результаты (W) в соответствующие клетки в скобках, рядом с первоначальными датами V, будем иметь следующую новую таблицу:

$M = 10$		$C_1 = 9$	$C_2 = 12$	$C_3 = 8$	$C_4 = 11$
$A_1 = 8$	$B_1 = 11$	7 (8)	9 (11)	8 (7)	12 (10)
	$B_2 = 9$	4 (6)	12 (9)	7 (5)	5 (8)
	$B_3 = 10$	10 (7)	9 (10)	3 (6)	10 (9)
$A_2 = 13$	$B_1 = 11$	11 (13)	19 (16)	14 (12)	12 (15)
	$B_2 = 9$	14 (11)	13 (14)	7 (10)	14 (13)
	$B_3 = 10$	11 (12)	13 (15)	12 (11)	16 (14)
$A_3 = 9$	$B_1 = 11$	12 (9)	11 (12)	5 (8)	12 (11)
	$B_2 = 9$	6 (7)	8 (10)	7 (6)	11 (9)
	$B_3 = 10$	6 (8)	14 (11)	9 (7)	7 (10)

Здесь для наглядности в заголовках показаны все ранее вычисленные частные средние A, B и C и общее среднее M.

Сравнивая в каждой клетке этой таблицы эмпирическую дату V с соответствующей ей теоретической датой W, получим значения всех случайных отклонений z по формуле:

$$z = V - M.$$

Проделав эти вычисления в действительности, найдем, что сумма всех этих отклонений:

$$\Sigma z = 0,$$

а сумма их квадратов:

$$\Sigma z^2 = 168.$$

Эта последняя сумма может служить числовой характеристикой колеблемости всех 36-ти дат V под влиянием „случайных" причин. По аналогии с ранее вычисленными суммами ($S_x = 450$, $S_a = 168$, $S_b = 24$ и $S_c = 90$) мы обозначим ее буквой S со значком z.

Итак у нас:

$$S_z = 168.$$

Сопоставляя эти суммы, легко убедиться в существовании следующего весьма важного равенства:

$$S_x = S_a + S_b + S_c + S_z,$$

показывающего, что общее рассеяние всех 36-ти дат V относительно их среднего M (измеряемое суммой S_x) состоит из отдельных компанентов (частей), зависящих только от определенных исследуемых факторов и от всей совокупности неустранимых „случайных" влияний.

В нашем примере:

$$450 = 168 + 24 + 90 + 168.$$

Здесь сумма $S_x = 450$ характеризует колеблемость дат под влиянием всех (как учтенных, так и не учтенных в данном опыте) факторов, сумма $S_a = 168$ измеряет колеблемость этих дат, вызываемую качественными различиями в отдельных вариантах фактора A (биологическими особенностями отдельных сортов или химическими особенностями отдельных агентов), сумма $S_b = 24$ дает представление о величине рассеяния этих же дат V под влиянием различий в вариантах второго фактора B (сроков посева, дозировок), сумма $S_c = 90$ количественно характеризует варьирование дат V, порождаемое различиями в вариантах третьего фактора C (почвенных особенностей, индивидуальных различий), а последняя сумма S_z измеряет силу всей совокупности неучтенных в этом опыте „случайных" факторов.

Последний компанент (S_z) нам особенно важен, так как он дает образец случайных (ничего не доказывающих) колебаний, на фоне сравнения с которыми мы и будем в дальнейшем оценивать значимость и существенность тех колебаний этих дат, которые свидетельствуют уже о закономерном влиянии на них со стороны тех или иных определенных изучаемых факторов.

III

При поверхностном рассмотрении всех этих сумм легко может показаться, что в данном опыте случайный фактор перекрывает большинство остальных (так как сумма $S_z = 168$ получилась здесь больше сумм $S_b = 24$ и $S_c = 90$), в силу чего у нас нет якобы особых оснований считать эффективность этих факторов „статистически доказанной". Такой вывод, однако, нельзя сделать путем простого сопоставления этих сумм, так как они с неодинаковой чуткостью отражают вариабельность отдельных факторов. Некоторые из этих сумм (напр.

$S_x = 450$), образуемые сравнительно большим количеством варьирующих элементов, подвержены вследствие этого меньшей колеблемости (под влиянием изменений каждого из этих элементов в отдельности), в то время как другие (напр. $S_b = 24$), зависящие от весьма малого числа таких колеблющихся элементов, даже при ничтожном изменении любого из них сразу же получают значительные сдвиги в сторону своего увеличения или уменьшения.

Подсчет общего количества таких вариагенных элементов приводит к понятию „числа степеней свободы", определяющих возникновение каждой из этих сумм. Так, в частности, сумма $S_x = 450$ характеризует варьирование 36-ти первоначальных дат V ($N = 36$) около их общего среднего $M = 10$. Все ли эти 36 варьирующих элементов колеблются здесь „свободно", т. е. при нашем их учете и изучении могут считаться принимающими какие угодно произвольные значения? Очевидно не все 36, а только 35 из них, так как одно последнее (все равно, какое именно) при таком исследовании должно быть искусственно подобрано с тем расчетом, чтобы среднее $M = 10$ при этом не изменилось. Это число (35) на единицу меньше общего количества всех варьирующих элементов ($N = 36$) и получает наименование „числа степеней свободы". Обозначив его буквой ν (ню), получим для нашей суммы S_x значение:

$$\nu^x = N - 1 = 36 - 1 = 35.$$

Вторая сумма $S_a = 168$ возникла в результате варьирования трех частных средних ($A_1 = 8$, $A_2 = 13$ и $A_3 = 9$) около того же среднего $M = 10$. По аналогичным соображениям из общего числа этих элементов ($a = 3$) „свободно" (т. е. не изменяя своего среднего $M = 10$) могут колебаться лишь 2 из них, откуда число степеней свободы:

$$\nu_a = a - 1 = 3 - 1 = 2.$$

Совершенно так же найдем, что число степеней свободы, определяющих возникновение третьей суммы $S_b = 24$ (при $b = 3$):

$$\nu_b = b - 1 = 3 - 1 = 2,$$

а в применении к сумме $S_c = 90$ (при $c = 4$):

$$\nu_c = C - 1 = 4 - 1 = 3.$$

Для определения числа степеней свободы последней суммы $S_z = 168$ удобно исходить из приведенного выше равенства:

$$S_x = S_a + S_b + S_c + S_z$$

Левая его часть (S_x) обладает числом степеней свободы $\nu_x = 35$, а в правой части для слагаемых S_a, S_b, S_c эти числа соответственно оказались равными $\nu_a = 2$, $\nu_b = 2$ и $\nu_c = 3$, что в сумме дает $2 + 2 + 3 = 7$. Следовательно, до полного числа 35 здесь недостает $35 - 7 = 28$ степеней свободы. Этот результат в дальнейшем мы будем получать непосредственно по формуле:

$$\nu_z = \nu_x - (\nu_a + \nu_b + \nu_c) = 35 - (2 + 2 + 3) = 35 - 7 = 28.$$

Для числовой характеристики варьирования всех наших дат V под действием тех или иных определенных факторов следует теперь каждую из вычисленных выше сумм S разделить на соответствующее число степеней свободы ν. В результате получим следующие меры варьирования (их квадраты), обозначаемые обычно буквой σ (сигма):

$$\sigma_x^2 = \frac{S_x}{\nu_x} = \frac{450}{35} = 12{,}86 \qquad \sigma_c^2 = \frac{S_c}{\nu_c} = \frac{90}{3} = 30$$

$$\sigma_a^2 = \frac{S_a}{\nu_a} = \frac{168}{2} = 64 \qquad \sigma_z^2 = \frac{S_z}{\nu_z} = \frac{168}{28} = 6$$

$$\sigma_b^2 = \frac{S_b}{\nu_b} = \frac{24}{2} = 12$$

Если бы мы пожелали охарактеризовать колеблемость наших дат под действием двух или нескольких факторов одновременно, то получили бы следующие значения соответствующих мер варьирования:

$$\sigma_{ab}^2 = \frac{S_a + S_b}{\nu_a + \nu_b} = \frac{168 + 24}{2 + 2} = \frac{192}{4} = 48$$

$$\sigma_{ac}^2 = \frac{S_a + S_c}{\nu_a + \nu_c} = \frac{168 + 90}{2 + 3} = \frac{258}{5} = 51{,}6$$

$$\sigma_{bc}^2 = \frac{S_b + S_c}{\nu_b + \nu_c} = \frac{24 + 90}{2 + 3} = \frac{114}{5} = 22{,}8$$

$$\sigma_{abc}^2 = \frac{S_a + S_b + S_c}{\nu_a + \nu_b + \nu_c} = \frac{168 + 24 + 90}{2 + 2 + 3} = \frac{282}{7} = 40{,}3$$

Итак, под влиянием совокупности чисто случайных воздействий наши даты V могли бы получить колебания около $M = 10$ с таким средним размахом, квадрат которого $\sigma_z^2 = 6$ (а стало быть $\sigma_z = \sqrt{6} = 2{,}45$). Под влиянием же действия отдельных изучаемых факторов A, B и C и их сочетаний квадраты этих средних размахов колебания наших дат V оказались значительно большими.

Отсюда намечается вывод о том, что отдельные варианты этих факторов имели здесь существенные качественные различия, которые мы не можем считать случайными. Для оценки достоверности этих выводов в применении к каждому отдельному фактору, посмотрим, во сколько раз квадрат характеризующей его меры варьирования превышает квадрат меры случайного варьирования $\sigma_z = 6$. Эти отношения мы обозначим буквами Θ (тэта) с соответствующими значками: a, b и c или их сочетаниями:

$$\Theta_a = \frac{\sigma_a^2}{\sigma_z^2} = \frac{84}{6} = 14$$

$$\Theta_b = \frac{\sigma_b^2}{\sigma_z^2} = \frac{12}{6} = 2$$

$$\Theta_c = \frac{\sigma_c^2}{\sigma_z^2} = \frac{30}{6} = 5$$

Итак, фактор A вызвал в нашем опыте настолько сильные колебания всех дат V, что квадрат меры этих колебаний ($\sigma_a^2 = 84$) превысил здесь квадрат меры случайного варьирования ($\sigma_z^2 = 6$) в 14 раз. Такое сильное превышение вряд ли может вызываться простой случайностью. В применении ко второму фактору (B) квадрат меры вызываемых им колебаний ($\sigma_b^2 = 12$) превысил квадрат меры случайного варьирования ($\sigma_z^2 = 6$) всего лишь в 2 раза, что повидимому все же оставляет некоторое сомнение в неслучайном происхождении этого превышения. Третий фактор (C) занимает в этом отношении промежуточное положение, так как квадрат меры зависящих от него колебаний ($\sigma_c^2 = 30$) превышает здесь квадрат той же меры случайного варьирования ($\sigma_z^2 = 6$) в 5 раз.

Для объективного решения вопроса о достоверности каждого из этих трех заключений обратимся к таблице Фишера-Стьюдента (см. в приложении).

IV

Для оценки по этой таблице достоверности показателя:

$$\Theta_a = \frac{\sigma_a^2}{\sigma_z^2} = \frac{84}{6} = 14$$

необходимо предварительно вспомнить, при каком числе степеней свободы был вычислен квадрат меры варьирования, стоящей в числителе (ν_1) и в знаменателе (ν_2). У нас здесь $\nu_1 = \nu_a = 2$, а $\nu_2 = \nu_z = 28$. Ищем в этой таблице вертикальный столбец, озаглавленный $\nu_1 = 2$ и горизонтальную строчку, соответствующую значению $\nu_z = 28$. На пересечении этого столбца и строчки находим клетку с такой записью:

3,34

5,45

8,93

Эти три числа обозначают те минимальные значения показателя Θ, которые при данных условиях (т. е. при $\nu_1 = 2$ и $\nu_z = 28$) могут условно считаться достоверными, гарантируя определенную вероятность справедливости такого заключения. В данной таблице использованы три последовательные значения этой вероятности: верхнее число (здесь 3,34) соответствует вероятности $P = 0{,}95$; среднее число (у нас 5,45,)—вероятности $P = 0{,}99$, а нижнее (8,93)—вероятности $P = 0{,}999$. Это надо понимать так: если бы оцениваемый квадрат меры варьирования σ_a^2 превысил квадрат меры случайного варьирования σ_z^2 в 3,34 раза ($\Theta = 3{,}34$), то мы могли бы утверждать, что при многократном повторении этого опыта при вполне аналогичных условиях в 95 случаях из 100 у нас получилось бы подтверждение данного вывода, а в оставшихся 5 случаях этот вывод мог бы оказаться и опровергнутым. Если значение Θ оказалось бы 5,45, то этот же вывод мы могли бы делать с 99 шансами (из 100), рискуя лишь в 1 случае принять кажущуюся закономерность за истинную. Наконец, если бы σ_a^2 превысила σ_z^2 в 8,93 раза, то такое значение Θ гарантировало бы справедливость нашего вывода в 999 случаях из 1000 (т. е. с риском ошибки в 1 случае на 1000). В нашем примере показатель $\Theta_a = 14$ явно превышает все три указанные в таблице его минимальные значения, в силу чего эффективность первого фактора A может считаться „статистически доказанной" (с вероятностью, превышающей $P = 0{,}999$).

В применении ко второму фактору:

$$\Theta_z = \frac{\sigma_b^2}{\sigma_z^2} = \frac{12}{6} = 2 \text{ (при } \nu_1 = 2 \text{ и } \nu_z = 28).$$

Это значение (2) оказывается ниже даже самого малого из 3-х чисел, стоящих в той же клетке таблицы Фишера-Стьюдента (3,34). Следовательно, значимость фактора B здесь статистически не доказана.

Для фактора C:

$$\Theta_c = \frac{\sigma_c^2}{\sigma_z^2} = \frac{30}{6} = 5 \text{ (при } \nu_1 = 3 \text{ и } \nu_2 = 28).$$

В соответствующей клетке таблицы Фишера-Стьюдента находим три числа: 2,95, 4,57 и 7,18. Данное значение $\Theta_c = 5$ превышает здесь первые два числа, но оказывается меньше последнего. Следовательно, значимость фактора C может считаться статистически доказанной с вероятностью, превышающей 0,99.

В дальнейшем мы будем условно обозначать найденную этим способом вероятность путем однократного (для $P = 0{,}95$), двукратного (для $P = 0{,}99$) или трехкратного (для $P = 0{,}999$) подчеркивания данного показателя Θ. В нашем примере следовательно:

$$\Theta_a = \underline{\underline{\underline{14}}} \qquad \Theta_b = 2 \qquad \Theta_c = \underline{\underline{5}}$$

Самые же вероятности 0,95, 0,99 и 0,999 мы будем для краткости обозначать символами P_5, P_1 и P_{01} (соответственно числу шансов из 100, могущих противоречить данному выводу).

Аналогично можно сопоставить с σ_z^2 тоже и меры σ_{ab}^2, σ_{ac}^2, σ_b^2 и σ_{abc}^2.

$$\Theta_{ab}^2 = \frac{\sigma_{ab}^2}{\sigma_z^2} = \frac{48}{6} = \underline{\underline{8}} \quad \text{(при } \nu_1 = 4 \text{ и } \nu_2 = 28)$$

$$\Theta_{ac}^2 = \frac{\sigma_{ac}^2}{\sigma_z^2} = \frac{51{,}6}{6} = \underline{\underline{\underline{8{,}6}}} \quad \text{(при } \nu_1 = 5 \text{ и } \nu_2 = 28)$$

$$\Theta_{bc}^2 = \frac{\sigma_{bc}^2}{\sigma_z^2} = \frac{22{,}8}{6} = \underline{\underline{3{,}8}} \quad \text{(при } \nu_1 = 5 \text{ и } \nu_2 = 28)$$

$$\Theta_{abc}^2 = \frac{\sigma_{abc}^2}{\sigma_z^2} = \frac{40{,}3}{6} = \underline{\underline{\underline{6{,}72}}} \quad \text{(при } \nu_1 = 7 \text{ и } \nu_2 = 28)$$

2

Там, где в таблице Фишера-Стьюдента не окажется столбца или строчки, соответствующих нужному числу степеней свободы, следует пользоваться столбцом или строчкой, соответствующими ближайшему меньшему числу (ν).

Для выяснения различий в эффективности отдельных факторов и их сочетаний, можно сравнить этим же способом квадраты соответствующих мер варьирования между собою (всюду бо̀льшую σ делим на меньшую):

$$\Theta_{a-b} = \frac{\sigma_a^2}{\sigma_b^2} = \frac{84}{12} = 7 \quad (\text{при } \nu_1 = 2 \text{ и } \nu_2 = 2)$$

$$\Theta_{a-c} = \frac{\sigma_a^2}{\sigma_c^2} = \frac{84}{30} = 2{,}8 \quad (\text{при } \nu_1 = 2 \text{ и } \nu_2 = 3)$$

$$\Theta_{c-b} = \frac{\sigma_c^2}{\sigma_b^2} = \frac{30}{12} = 2{,}5 \quad (\text{при } \nu_1 = 3 \text{ и } \nu_2 = 2)$$

$$\Theta_{ac-ab} = \frac{\sigma_{ac}^2}{\sigma_{ab}^2} = \frac{51{,}6}{48} = 1{,}75 \quad (\text{при } \nu_1 = 5 \text{ и } \nu_2 = 4)$$

$$\Theta_{ab-bc} = \frac{\sigma_{ab}^2}{\sigma_{bc}^2} = \frac{48}{22{,}8} = 2{,}11 \quad (\text{при } \nu_1 = 4 \text{ и } \nu_2 = 5)$$

$$\Theta_{ac-bc} = \frac{\sigma_{ac}^2}{\sigma_{bc}^2} = \frac{51{,}6}{22{,}8} = 2{,}26 \quad (\text{при } \nu_1 = 5 \text{ и } \nu_2 = 5)$$

В нашем примере все эти различия остаются недоказанными (бо̀льший эффект действия одних факторов или их сочетаний по сравнению с другими может вызываться здесь также и простой случайностью).

На этом заканчивается первая фаза математико-статистической обработки комплекса признаков. В применении к схеме полевого опыта по отношению к отдельным факторам мы должны были бы, следовательно, утверждать, что на величину урожая V здесь несомненно влияют различия в сортах (A) и в почвенных особенностях на отдельных участках поля (C). Что же касается сроков посева (B), то влияние их на урожай V здесь не может считаться установленным.

В случае же физиологической иллюстрации этого примера следовало бы посчитать установленным эффективность химических различий отдельных агентов (A) и значимость индивидуальных особенностей четырех объектов этого опыта (C); испытанные же дозировки этих агентов (B) пришлось бы из осторожности признать нейтральными.

V

Вторая фаза статистического анализа этих данных заключается в выяснении различий между отдельными вариантами тех факторов, влияние которых на результативный признак V было статистически доказано в первой фазе этой разработки (у нас, следовательно, в применении лишь к факторам A и C) за вычетом тех факторов, которые были введены в опыт лишь в целях его уточнения, сами же по себе для экспериментатора особого интереса не представляют (напр., почвенные особенности отдельных участков опытного поля или индивидуальные свойства отдельных объектов испытания, т. е. у нас фактор C).

Вспомним, что группа частных средних, характеризующих отдельные варианты фактора A, у нас оказалась состоящей из трех ($a = 3$) средних:

$$A_1 = 8 \qquad A_2 = 13 \qquad A_3 = 9.$$

Для сопоставления их между собой и выяснения достоверности (существенности, неслучайности, объективной убедительности) всех этих различий необходимо вычислить средние ошибки m (их квадраты m^2), характеризующие средние пределы ожидаемых (возможных) чисто случайных колебаний всех этих частных средних. Вот формула для их вычисления:

$$m_a^2 = \frac{a \cdot \sigma_z^2}{N} = \frac{3 \cdot 6}{36} = \frac{18}{36} = 0,5$$

В применении к трем частным средним второго фактора B (если бы нам понадобилось их анализировать):

$$m_b^2 = \frac{b \cdot \sigma_z^2}{N} = \frac{3 \cdot 6}{36} = \frac{18}{36} = 0,5,$$

а в применении к четырем частным средним фактора C:

$$m_c^2 = \frac{c \cdot \sigma_z^2}{N} = \frac{4 \cdot 6}{36} = \frac{24}{36} = 0,67.$$

Так как нам придется сравнивать эти частные средние попарно (в пределах каждого отдельного фактора), то для оценки достоверности этих разностей необходимо вычислить ее среднюю ошибку Δ (дельта) по формуле:

$$\Delta = \sqrt{2m^2}$$

Так, в частности, для фактора A:

$$\Delta_a = \sqrt{2m_a^2} = \sqrt{2 \cdot 0,5} = \sqrt{1} = 1.$$

В применении к фактору B:

$$\Delta_b = \sqrt{2m_b^2} = \sqrt{2 \cdot 0,5} = \sqrt{1} = 1.$$

В применении к фактору C:

$$\Delta_c = \sqrt{2m_c^2} = \sqrt{2 \cdot 0,67} = \sqrt{1,34} = 1,16.$$

Для оценки достоверности отдельных различий между частными средними A_1, A_2 и A_3 выясним, во сколько раз каждая оцениваемая их разность превышает соответствующую среднюю ошибку этой разности (Δ). Так, в частности, сопоставляя средние $A_2 = 13$ и $A_1 = 8$, мы находим, что разность их $A_2 - A_1 = 13 - 8 = 5$ превышает здесь среднюю ошибку этой разности $\Delta_a = 1$ в 5 раз. Обозначив это отношение буквою t, получим:

$$t = \frac{A_2 - A_1}{\Delta_a} = \frac{13 - 8}{1} = \frac{5}{1} = 5.$$

Аналогично найдем, что разность между $A_2 = 13$ и $A_3 = 9$ превышает свою среднюю ошибку (ту же самую) Δ_a в 4 раза:

$$t = \frac{A_2 - A_3}{\Delta_a} = \frac{13 - 9}{1} = \frac{4}{1} = 4.$$

В применении же к разности между $A_3 = 9$ и $A_1 = 8$ это отношение:

$$t = \frac{A_3 - A_1}{\Delta_a} = \frac{9 - 8}{1} = \frac{1}{1} = 1.$$

Для оценки достоверности всех этих различий обратимся к той же таблице Фишера-Стьюдента, где в крайнем справа столбце (озаглавленном сверху буквой t) следует отыскать значение t, соответствующее числу степеней свободы:

$$\nu = \nu_z = 28.$$

В соответствующей клетке там стоят 3 числа:

2,05
2,76
3,67

Первое из них (2,05) представляет собой такое значение t, которое гарантирует вероятность $P_5 = 0{,}95$ (будем обозначать его t_5); второе (2,76) соответствует вероятности $P_1 = 0{,}99$ (обозначим его t_1), а третье (3,67)—вероятности $P_{01} = 0{,}999$ t_{01}.

Первые два из трех полученных выше действительных значений t (у нас 5 и 4) превысили здесь эти минимальные числа таблицы, а последнее ($t = 1$) оказалось заметно меньше их. Следовательно, первые два различия (между A_2 и A_1 и между A_2 и A_3) можно считать статистически доказанными (с вероятностью, превышающей $P_{01} = 0{,}999$), а последнюю разницу этих частных средних ($A_3 - A_1$) из осторожности придется здесь посчитать лишь случайной:

$$\underline{\underline{t = 5}} \qquad \underline{\underline{t = 4}} \qquad t = 1.$$

В дальнейшем мы не будем вычислять этих значений t, а ограничимся определением лишь минимальных значений достоверных разностей между отдельными парами частных средних (обозначим их буквами l) по формуле:

$$l = \Delta t.$$

В нашем примере:

$$\Delta_a = 1 \qquad t_5 = 2{,}00 \qquad t_1 = 2{,}66 \qquad t_{01} = 3{,}36.$$

По этим денным вычислим три значения l:

$$l_5 = \Delta_a t_5 = 1 \cdot 2{,}05 = 2{,}05$$
$$l_1 = \Delta_a t_1 = 1 \cdot 2{,}76 = 2{,}76$$
$$l_{01} = \Delta_a t_{01} = 1 \cdot 3{,}67 = 3{,}67.$$

В нашем примере разности $A_2 - A_1 = 13 - 8 = 5$ и $A_2 - A_3 = 13 - 9 = 4$ оказались больше всех этих значений t, следовательно, все они свидетельствуют о достоверных (не случайных) различиях между соответствующими вариантами фактора A; разность же $A_2 - A_3 = 9 - 8 = 1$ этому условию очевидно не удовлетворяет. Если бы какая-нибудь разность оказалась больше только первого значения $t_5 = 2{,}05$, то мы считали бы ее статистически доказанной с вероятностью, превышающей $P_5 = 0{,}95$. Если бы она превысила также и следующее значение $t_1 = 2{,}76$, мы могли бы утверждать то же положение с вероятностью $P_1 = 0{,}99$. В нашем же примере первые две разности (5 и 4), превысившие третье значение $t_{01} = 3{,}67$, свидетельствуют о существенности этих различий с вероятностью, превышающей $P_{01} = 0{,}999$.

Аналогичную оценку достоверности различий между отдельными вариантами можно было бы произвести и в применении к остальным двум факторам (B и C). Однако в нашем примере этого можно и не делать, так как достоверность фактора B (в целом) в первой фазе этого анализа осталась недоказанной, а последний фактор (C) введен был в опыт лишь для его уточнения, и отдельные варианты этого фактора для исследователя самостоятельного интереса не представляют.

VI

Предположим теперь, что все три изучаемые в данном опыте фактора (A, B и C) представляют собой для исследователя одинаковый интерес (например, фактор C в схеме полевого опыта характеризует не какие-либо произвольные участки поля, а определенные группы делянок с различным удобрением почвы или, при физиологической иллюстрации, этот фактор C относится не к четырем случайным индивидуальностям, а к четырем представителям каких-либо определенных изучаемых типов). В этом случае экспериментатора может заинтересовать также и совместное действие нескольких факторов (например, A и B, A и C, B и C, а в более сложных схемах также и различные сочетания по 3 и более факторов одновременно). Пусть, например, требуется выяснить совместное действие на V факторов A и C.

Составляем новую комбинационную таблицу, в клетки которой заносим средние из отдельных дат V, относящиеся к каждому сочетанию всех вариантов фактора A с вариантами фактора C:

	C_1	C_2	C_3	C_4
A_1	7	10	6	9
A_2	12	15	11	14
A_3	8	11	7	10

Все эти числа получены как средние из соответствующих дат (V) нашей первоначальной таблицы, например:

$$\frac{7+4+10}{3}=\frac{21}{3}=7 \qquad \frac{9+12+9}{3}=\frac{30}{3}=10 \text{ и т. д.}$$

Для выяснения вероятности различий между отдельными средними, помещенными в этой таблице, необходимо вычислить предварительно их общую среднюю ошибку по формуле:

$$m_{ac}^2=\frac{ac\sigma_z^2}{N}=\frac{3\cdot 4\cdot 6}{36}=\frac{72}{36}=2.$$

Для средних, сгруппированных по отдельным сочетаниям факторов A и B, эту ошибку следовало бы вычислить по формуле:

$$m_{ab}^2=\frac{ab\sigma_z^2}{N}=\frac{3\cdot 3\cdot 6}{36}=\frac{54}{36}=1{,}5,$$

а в применении к комбинациям факторов B и C — по формуле:

$$m_{bc}^2=\frac{bc\sigma_z^2}{N}=\frac{3\cdot 4\cdot 6}{36}=\frac{72}{36}=2.$$

Средняя ошибка разности этих средних в рассматриваемом случае:

$$\Delta_{ac}=\sqrt{2m_{ac}^2}=\sqrt{2\cdot 2}=\sqrt{4}=2,$$

а в остальных двух:

$$\Delta_{ab}=\sqrt{2m_{ab}^2}=\sqrt{2\cdot 1{,}5}=\sqrt{3}=1{,}73$$

$$\Delta_{bc}=\sqrt{2m_{bc}^2}=\sqrt{2\cdot 2}=\sqrt{4}=2.$$

Вычисление минимальных разностей l (для $\Delta_{ac}=2$ и при $\nu=\nu_z=28$) дает:

$$l_5=\Delta_{ac}t_5=2\cdot 2{,}05=4{,}10$$

$$l_1=\Delta_{ac}t_1=2\cdot 2{,}76=5{,}52$$

$$l_{01}=\Delta_{ac}\,t_{a1}=2\cdot 3{,}67=7{,}34.$$

Сопоставляя отдельные фактически полученные разности наших средних (см. последнюю таблицу) с этими тремя минимальными разностями l, можно судить об их достоверности. Так, например, сочетание варианта A_2 с вариантом C_2, давшее среднюю величину результативного признака $=15$, по сравнению с эффективностью комбинации A_1 и C_3, давшей средний результат $=6$, является несомненно более выгодным, так как разность $15-6=9$ явно превышает здесь наибольшее значение $l_{01}=7{,}34$. Комбинация же A_1C_1 (давшая 7) по сравнению с комбинацией A_3C_4 (давшей 10) не является здесь существенно менее выгодной, так как разность $10-7=3$ оказывается меньше даже самого малого значения $l_5=4{,}10$. Сравнение сочетаний A_2C_4 (14) и A_3C_1 (8) дает право считать соответствующее различие статистически доказанным с вероятностью, превышающей 0,99 (так как $14-8=6$ больше $l_1=5{,}52$).

Аналогично можно было бы проанализировать и другие сочетания этих факторов (например, A и B, B и C).

[illegible] $N = 10$ [illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

[illegible]

[illegible]

		C_1		C_2		C_3		C_4		
		D_1	D_2	D_1	D_2	D_1	D_2	D_1	D_2	
A_1	B_1	6 (36)	2 (4)	4 (16)	0 (0)	7 (49)	3 (9)	5 (25)	1 (1)	28
	B_2	3 (9)	3 (9)	1 (1)	1 (1)	4 (16)	4 (16)	2 (4)	2 (4)	20
A_2	B_1	5 (25)	7 (49)	3 (9)	5 (25)	6 (36)	8 (64)	4 (16)	6 (36)	44
	B_2	8 (64)	2 (4)	6 (36)	0 (0)	9 (81)	3 (9)	7 (49)	1 (1)	36
A_3	B_1	7 (49)	3 (9)	5 (25)	1 (1)	8 (64)	4 (16	6 (36)	2 (4)	36
	B_2	4 (16)	4 (16)	2 (4)	2 (4)	5 (25)	5 (25)	3 (9)	3 (9)	28
		33	21	21	9	39	27	27	15	192

$a=3 \quad b=2 \quad c=4 \quad d=2 \quad N=abcd=3\cdot 2\cdot 4\cdot 2=48$

$\Sigma V=192 \qquad \Sigma V^2=1\,016$

2) Вычислить частные средние $A, B, C, D\ldots$ характеризующие отдельные варианты всех факторов и общее среднее M для всего опыта в целом:

$$\begin{cases} A_1=\dfrac{28+20}{16}=\dfrac{48}{16}=3 \\ A_2=\dfrac{44+36}{16}=\dfrac{80}{16}=5 \\ A_3=\dfrac{36+28}{16}=\dfrac{64}{16}=4 \end{cases} \qquad \begin{cases} C_1=\dfrac{33+21}{12}=\dfrac{54}{12}=4{,}5 \\ C_2=\dfrac{21+9}{12}=\dfrac{30}{12}=2{,}5 \\ C_3=\dfrac{39+27}{12}=\dfrac{66}{12}=5{,}5 \\ C_4=\dfrac{27+15}{12}=\dfrac{42}{12}=3{,}5 \end{cases}$$

$$\begin{cases} B_1=\dfrac{28+44+36}{24}=\dfrac{108}{24}=4{,}5 \\ B_2=\dfrac{20+36+28}{24}=\dfrac{84}{24}=3{,}5 \end{cases} \qquad \begin{cases} D_1=\dfrac{33+21+39+27}{24}=\dfrac{120}{24}=5 \\ D_2=\dfrac{21+9+27+15}{24}=\dfrac{72}{24}=3 \end{cases}$$

$$M=\frac{\Sigma V}{N}=\frac{192}{48}=4$$

3) Вычислить суммы квадратов отклонений $\alpha, \beta, \gamma, \delta\ldots$ этих частных средних $A, B, C, D\ldots$ от общего среднего M:

A	α	α^2
3	-1	1
5	$+1$	1
4	0	0
	0	2

B	β	β^2
4,5	$+0{,}5$	0,25
3,5	$-0{,}5$	0,25
	0	0,5

C	γ	γ^2
4,5	$+0{,}5$	0,25
2,5	$-1{,}5$	2,25
5,5	$+1{,}5$	2,25
3,5	$-0{,}5$	0,25
	0	5

D	δ	δ^2
5	$+1$	1
3	-1	1
	0	2

$\Sigma\alpha^2=2 \qquad \Sigma\beta^2=0{,}5 \qquad \Sigma\gamma^2=5 \qquad \Sigma\delta^2=2$

4) Вычислить суммы S_x, S_a, S_b, S_c, S_d и S_z, характеризующие общее рассеяние всех дат V и отдельные компаненты этого рассеяния, зависящие только от действия изучаемых факторов A, B, C, D и от всей совокупности неучтенных в данном опыте („случайных") факторов (сумма S_z), по формулам:

$$S_x = \Sigma V^2 - N(M)^2 = 1016 - 48(4)^2 = 1016 - 48 \cdot 16 = 1016 - 768 =$$
$$= 248 \text{ (новый прием)}$$

$$S_a = \frac{N\Sigma\alpha^2}{a} = \frac{48 \cdot 2}{3} = 32 \qquad S_c = \frac{N\Sigma\gamma^2}{c} = \frac{48 \cdot 5}{4} = 60$$

$$S_b = \frac{N\Sigma\beta^2}{b} = \frac{48 \cdot 0{,}5}{2} = 12 \qquad S_d = \frac{N\Sigma\delta^2}{d} = \frac{48 \cdot 2}{2} = 48$$

$$S_z = S_x - (S_a + S_b + S_c + S_d) = 248 - (32 + 12 + 60 + 48) =$$
$$= 248 - 152 = 96.$$

5) Определить „число степеней свободы" (ν) для каждой из этих сумм по формулам:

$$\nu_a = a - 1 = 3 - 1 = 2$$
$$\nu_b = b - 1 = 2 - 1 = 1$$
$$\nu_c = c - 1 = 4 - 1 = 3$$
$$\nu_d = d - 1 = 2 - 1 = 1$$

$$\nu_x = N - 1 = 48 - 1 = 47$$
$$\nu_z = \nu_x - (\nu_a + \nu_b + \nu_c + \nu_d) =$$
$$= 47 - (2 + 1 + 3 + 1) = 47 - 7 = 40.$$

6) Вычислить квадраты мер варьирования по формулам:

$$\sigma_a^2 = \frac{S_a}{\nu_a} = \frac{32}{2} = 16 \qquad \sigma_c^2 = \frac{S_c}{\nu_c} = \frac{60}{3} = 20$$

$$\sigma_b^2 = \frac{S_b}{\nu_b} = \frac{12}{1} = 12 \qquad \sigma_d^2 = \frac{S_d}{\nu_d} = \frac{48}{1} = 48$$

$$\sigma_x^2 = \frac{S_x}{\nu_x} = \frac{248}{47} = 5{,}28 \qquad \sigma_z^2 = \frac{S_z}{\nu_z} = \frac{96}{40} = 2{,}4,$$

а при желании и квадраты мер варьирования по совокупности нескольких признаков, например, для следующих заинтересовавших исследователя комбинаций:

$$\sigma^2_{ab} = \frac{S_a + S_b}{\nu_a + \nu_b} = \frac{32 + 12}{2 + 1} = \frac{44}{3} = 14{,}67$$

$$\sigma^2_{ad} = \frac{S_a + S_d}{\nu_a + \nu_d} = \frac{32 + 48}{2 + 1} = \frac{80}{3} = 26{,}57$$

$$\sigma^2_{bc} = \frac{S_b + S_c}{\nu_b + \nu_c} = \frac{12 + 60}{1 + 3} = \frac{72}{4} = 18{,}00$$

$$\sigma^2_{abc} = \frac{S_a + S_b + S_c}{\nu_a + \nu_b + \nu_c} = \frac{32 + 12 + 60}{2 + 1 + 3} = \frac{104}{6} = 17{,}33$$

$$\sigma^2_{bcd} = \frac{S_b + S_c + S_d}{\nu_b + \nu_c + \nu_d} = \frac{12 + 60 + 48}{1 + 3 + 1} = \frac{120}{5} = 24{,}00$$

$$\sigma^2_{abcd} = \frac{S_a + S_b + S_c + S_d}{\nu_a + \nu_b + \nu_c + \nu_d} = \frac{32 + 12 + 60 + 48}{2 + 1 + 3 + 1} = \frac{152}{7} = 21{,}71.$$

7) Вычислить отношения Θ, показывающие, во сколько раз данный квадрат меры варьирования превышает квадрат меры случайного варьирования (σ_z^2), и оценить достоверность этого превышения по таблице Фишера-Стьюдента при числе степеней свободы ν_1 (для σ, стоящей в числителе) и ν_2 (для σ, стоящей в знаменателе) по формулам:

$$\Theta_a = \frac{\sigma_a^2}{\sigma_z^2} = \frac{16}{2{,}4} = \underline{\underline{6{,}67}} \text{ (при } \nu_1 = 2 \text{ и } \nu_2 = 40)$$

$$\Theta_b = \frac{\sigma_b^2}{\sigma_z^2} = \frac{12}{2{,}4} = \underline{5{,}00} \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 40)$$

$$\Theta_c = \frac{\sigma_c^2}{\sigma_z^2} = \frac{20}{2{,}4} = \underline{\underline{\underline{8{,}33}}} \text{ (при } \nu_1 = 3 \text{ и } \nu_2 = 40)$$

$$\Theta_d = \frac{\sigma_d^2}{\sigma_z^2} = \frac{48}{2{,}4} = \underline{\underline{\underline{20{,}00}}} \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 40,$$

а при желании также и в применении к мерам варьирования, характеризующим совокупное действие этих факторов, например:

$$\Theta_{bc} = \frac{\sigma^2_{bc}}{\sigma^2_z} = \frac{18}{2{,}4} = 7{,}5 \text{ (при } \nu_1 = 4 \text{ и } \nu_2 = 40)$$

$$\Theta_{bcd} = \frac{\sigma^2_{bcd}}{\sigma^2_z} = \frac{24}{2{,}4} = 10 \text{ (при } \nu_1 = 5\ \nu_2 = 40) \text{ и т. д.}$$

(за неимением в таблице подходящего столбца или строчки, производим оценку по месту, соответствующему ближайшему меньшему числу степеней свободы).

8) Вычислить аналогичные показатели Θ для сравнения эффективности отдельных факторов (всюду бо́льшую σ делить на меньшую):

$$\Theta_{a-b} = \frac{\sigma_a^2}{\sigma_b^2} = \frac{16}{12} = 1{,}33 \text{ (при } \nu_1 = 2 \text{ и } \nu_2 = 1)$$

$$\Theta_{c-a} = \frac{\sigma_c^2}{\sigma_a^2} = \frac{20}{16} = 1{,}25 \text{ (при } \nu_1 = 3 \text{ и } \nu_2 = 2)$$

$$\Theta_{d-a} = \frac{\sigma_d^2}{\sigma_a^2} = \frac{48}{16} = 3{,}00 \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 2)$$

$$\Theta_{c-b} = \frac{\sigma_c^2}{\sigma_b^2} = \frac{20}{12} = 1{,}67 \text{ (при } \nu_1 = 3 \text{ и } \nu_2 = 1)$$

$$\Theta_{d-b} = \frac{\sigma_d^2}{\sigma_b^2} = \frac{48}{12} = 4{,}00 \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 1)$$

$$\Theta_{d-c} = \frac{\sigma_d^2}{\sigma_c^2} = \frac{48}{20} = 2{,}40 \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 3)$$

(достоверных показателей здесь не оказалось вовсе), а при желании оценить достоверность различия в эффективности также и двух каких-либо комбинаций этих факторов, например:

$$\Theta_{ad-bc} = \frac{\sigma^2_{ad}}{\sigma^2_{bc}} = \frac{26{,}67}{18{,}00} = 1{,}48 \text{ (при } \nu_1 = 3 \text{ и } \nu_2 = 4)$$

(не достоверно).

9) Вычислить квадраты средних ошибок (m), характеризующих средние пределы случайных колебаний отдельных частных средних A, B, C, D... по формулам:

$$m_a^2 = \frac{a\sigma_z^2}{N} = \frac{3 \cdot 2{,}4}{48} = 0{,}15 \qquad m_c^2 = \frac{c\sigma_z^2}{N} = \frac{4 \cdot 2{,}4}{48} = 0{,}20$$

$$m_b^2 = \frac{b\sigma_z^2}{N} = \frac{2 \cdot 2{,}4}{48} = 0{,}10 \qquad m_d^2 = \frac{d\sigma_z^2}{N} = \frac{2 \cdot 2{,}4}{48} = 0{,}10.$$

10) Вычислить средние ошибки разностей отдельных частных средних (при сопоставлении их попарно в пределах каждого фактора) по формулам:

$$\Delta_a = \sqrt{2m_a^2} = \sqrt{2 \cdot 0{,}15} = \sqrt{0{,}30} = 0{,}55$$

$$\Delta_b = \sqrt{2m_b^2} = \sqrt{2 \cdot 0{,}10} = \sqrt{0{,}20} = 0{,}45$$

$$\Delta_c = \sqrt{2m_c^2} = \sqrt{2 \cdot 0{,}20} = \sqrt{0{,}40} = 0{,}63$$

$$\Delta_d = \sqrt{2m_d^2} = \sqrt{2 \cdot 0{,}10} = \sqrt{0{,}20} = 0{,}45.$$

11) Вычислить минимальные разности l_5, l_1 и l_{01}, наличие которых гарантирует существенность соответствующих различий между парами частных средних с вероятностью $P_5 = 0{,}95$, $P_1 = 0{,}99$ и $P_{01} = 0{,}999$ по формулам типа: $l = \Delta t$, где t находится по таблице Фишера-Стьюдента при числе степеней свободы $\nu = \nu_z$.

У нас $\nu = 40$ или ближайшее меньшее $\nu = 30$. Следовательно, $t_5 = 2{,}04$, $t_1 = 2{,}75$ и $t_{01} = 3{,}65$.

Для фактора A

$$l_5 = \Delta_a t_5 = 0{,}55 \cdot 2{,}04 = 1{,}12$$

$$l_1 = \Delta_a t_1 = 0{,}55 \cdot 2{,}75 = 1{,}51$$

$$l_{01} = \Delta_a t_{01} = 0{,}55 \cdot 3{,}65 = 2{,}01$$

Для фактора B

$$l_5 = \Delta_b t_5 = 0{,}45 \cdot 2{,}04 = 0{,}92$$

$$l_1 = \Delta_b t_1 = 0{,}45 \cdot 2{,}75 = 1{,}24$$

$$l_{01} = \Delta_b t_{01} = 0{,}45 \cdot 3{,}65 = 1{,}64$$

Для фактора C

$$l_5 = \Delta_c t_5 = 0{,}63 \cdot 2{,}04 = 1{,}29$$

$$l_1 = \Delta_c t_1 = 0{,}63 \cdot 2{,}75 = 1{,}73$$

$$l_{01} = \Delta_c t_{01} = 0{,}63 \cdot 3{,}65 = 2{,}30$$

Для фактора D

$l_5 = \Delta$ [illegible]

[illegible]

а для комбинаций трех факторов A, B и C:

		C_1	C_2	C_3	C_4
A_1	B_1	4	2	5	3
	B_2	3	1	4	2
A_2	B_1	6	4	7	5
	B_2	5	3	6	4
A_3	B_1	5	3	6	4
	B_2	4	2	5	3

Аналогично можно было бы получить и все другие таблицы (по мере фактической надобности).

13) Вычислить квадраты средних ошибок m по следующим формулам (используются лишь по мере надобности для отдельных фактически составленных таблиц):

$$m^2_{ab} = \frac{ab\sigma_z^2}{N} \qquad m^2_{abc} = \frac{abc\sigma_z^2}{N}$$

$$m^2_{ac} = \frac{ac\sigma_z^2}{N} \qquad m^2_{abd} = \frac{abd\sigma_z^2}{N}$$

$$m^2_{ad} = \frac{ad\sigma_z^2}{N} \qquad m^2_{acd} = \frac{acd\sigma_z^2}{N}$$

$$m^2_{bc} = \frac{bc\sigma_z^2}{N} \qquad m^2_{bcd} = \frac{bcd\sigma_z^2}{N}$$

$$m^2_{bd} = \frac{bd\sigma_z^2}{N} \qquad m^2_{abcd} = \frac{abcd\sigma_z^2}{N}$$

$$m^2_{cd} = \frac{cd\sigma_z^2}{N}$$

В нашем примере для таблицы AC:

$$m^2_{ac} = \frac{3 \cdot 4 \cdot 2{,}4}{48} = 0{,}60.$$

14) Для сравнения между собой отдельных объединенных (средних) дат, внутри каждой таблицы вычислить средние ошибки разностей Δ и три минимальные разности l по обычным формулам (см. № 10 и 11).

В нашем примере для таблицы AC (при числе степеней свободы $\nu = \nu_z = 40$ или ближайшее меньшее $\nu = 30$):

$$\Delta_{ac} = \sqrt{2m^2_{ac}} = \sqrt{2 \cdot 0{,}60} = \sqrt{1{,}20} = 1{,}10$$

$$l_5 = \Delta_{ac} t_5 = 1{,}10 \cdot 2{,}04 = 2{,}24$$

$$l_1 = \Delta_{ac} t_1 = 1{,}10 \cdot 2{,}75 = 3{,}03$$

$$l_{01} = \Delta_{ac} t_{01} = 1{,}10 \cdot 3{,}65 = 4{,}02$$

Выше мы рассмотрели классический случай абсолютно строгого соблюдения поставленных ранее требований: полного (без пропусков) заполнения датами V всех N клеток комбинационной таблицы и использования всех без исключения мыслимых сочетаний вариантов отдельных факторов (взятых как по два, так и по нескольку) и притом совершенно одинаковое число раз.

Покажем приемы статистической обработки комплекса признаков, невполне удовлетворяющего этим условиям. Вот иллюстрирующая эту схему комбинационная таблица:

	B_1		B_2		B_3		B_4	
	№ 1 $C_1D_1(E_1)$		№ 5 $C_2D_2(E_2)$		№ 9 $C_3D_3(E_3)$		№ 13 $C_4D_4(E_4)$	
A_1	7	7	10	9	10	9	13	15
	№ 2 $C_2D_4(E_3)$		№ 6 $C_1D_3(E_4)$		№ 10 $C_4D_2(E_1)$		№ 14 $C_3D_1(E_2)$	
A_2	9	8	11	7	W_2	10	1	0
	№ 3 $C_3D_2(E_4)$		№ 7 $C_4D_1(E_3)$		№ 11 $C_1D_4(E_2)$		№ 15 $C_2D_3(E_1)$	
A_3	1	3	W_1	1	8	11	3	3
	№ 4 $C_4D_3(E_2)$		№ 8 $C_3D_4(E_1)$		№ 12 $C_2D_1(E_4)$		№ 16 $C_1D_2(E_3)$	
A_4	3	2	4	4	2	4	3	2

Здесь испытываются 4 фактора (A, B, C и D), каждый из которых встречается в одинаковом числе вариантов (всюду по 4). Последнее условие является здесь обязательным. Следовательно, у нас:

$$a = b = c = d = 4\ (= h).$$

Общее число испытанных здесь комбинаций (отдельных клеток таблицы) равно всего лишь:

$$4 \times 4 = 16,$$

а не $abcd = 256$, какое потребовалось бы обеспечить, чтобы испытать все сочетания вариантов при совместном действии двух, трех и четырех факторов одновременно. Следовательно, здесь исследуется лишь изолированное действие отдельных факторов (A, B, C и D) и их совместное действие по два (A и B, A и C, A и D, B и C, B и D, C и D). Исследовать же эффективность тройных (и выше) комбинаций здесь нельзя.

В таблице варианты фактора A указаны слева, варианты фактора B — сверху, а варианты факторов C и D обозначены соответствующими символами внутри отдельных клеток этой таблицы, после порядкового номера, стоящего в левом верхнем углу. Пятый фактор E, фактически здесь не использованный, указан соответствующими буквами E, поставленными в скобках. В каждой такой клетке записаны по две даты V, соответствующие повторному (двукратному) измерению (или наблюдению) результативного признака V при вполне аналогичных условиях (т. е. при той же самой комбинации вариантов A, B, C и D). Таким образом у нас общее число запроектированных дат $N=32$.

Предположим однако, что не все они были здесь получены в действительности, и некоторые выпали (опыт не удался). На соответствующих местах (в клетках за №№ 7 и 10) в этой таблице поставлены обозначения „теоретических дат": W_1 и W_2, которыми мы в дальнейшем заменим эти отсутствующие „эмпирические даты" V_1 и V_2.

Легко убедиться в том, что в 16 клетках нашей таблицы использованы все мыслимые парные сочетания отдельных вариантов каждого фактора с вариантами всех остальных.

(Такие табличные схемы получили наименования „греко-латинских квадратов" в соответствии с общепринятым ранее способом их изображения при помощи букв латинского и греческого алфавитов).

Для вычисления частных средних, соответствующих четырем вариантам фактора A, следует сложить даты V по отдельным строчкам таблицы и каждую такую сумму разделить на 8. При этом в тех строчках, в которых окажутся выпадающие даты, мы к полученным фактически числовым суммам припишем недостающие слагаемые в виде соответствующих буквенных обозначений (W_1 или W_2), а самое вычисление этих средних пока не доведем до конца:

$$A_1=\frac{7+7+10+9+10+9+13+15}{8}=\frac{80}{8}=10$$

$$A_2=\frac{9+8+11+7+W_2+10+1+0}{8}=\frac{46+W_2}{8} \text{ (не закончено)}$$

$$A_3=\frac{1+3+W_1+1+8+11+3+3}{8}=\frac{30+W_1}{8} \text{ (не закончено)}$$

$$A_4=\frac{3+2+4+4+2+4+3+2}{8}=\frac{24}{8}=3$$

При вычислении группы частных средних по признаку B будем суммировать даты по соответствующим вертикальным столбцам:

$$B_1=\frac{7+7+9+8+1+3+3+2}{8}=\frac{40}{8}=5$$

$$B_2=\frac{9+10+11+7+W_1+1+4+4}{8}=\frac{46+W_1}{8} \text{ (не закончено)}$$

$$B_3=\frac{10+9+W_2+10+8+11+2+4}{8}=\frac{54+W_2}{8} \text{ (не закончено)}$$

$$B_4=\frac{13+15+1+0+3+3+3+2}{8}=\frac{40}{8}=5.$$

В нашем примере число факторов $f=4$, следовательно:

$$W=A+B+C+D-3M.$$

Теоретическая дата W_1 относится к той клетке нашей таблицы (№ 7), в которой записаны результаты совместного действия следующих вариантов отдельных факторов:

$$A_3 \qquad B_2 \qquad C_4 \quad \text{и} \quad D_1$$

Подставляя соответствующие значки в наше последнее уравнение, получим:

$$W_3=A_1+B_2+C_4+D_1-3M.$$

Вспомним, что эти четыре частные средние и общее среднее M у нас оказались при вычислении неокончательными, имея следующий вид:

$$A_3=\frac{30+W_1}{8} \qquad C_4=\frac{44+W_1+W_2}{8}$$

$$B_2=\frac{46+W_1}{8} \qquad D_1=\frac{22+W_1}{8},$$

а общее среднее $M=\frac{180+W_1+W_2}{32}$.

Подставляя эти данные в наше уравнение, получим:

$$W_1=\frac{30+W_1}{8}+\frac{46+W_1}{8}+\frac{44+W_1+W_2}{8}+$$

$$+\frac{22+W_1}{8}-3\left(\frac{180+W_1+W_2}{32}\right)$$

или, после соответствующих преобразований:

$$19W_1-W_2=28.$$

Вторая теоретическая дата (в клетке № 10) обусловлена действием следующей комбинации вариантов отдельных факторов:

$$A_2 \qquad B_3 \qquad C_4 \quad \text{и} \quad D_2$$

следовательно:

$$W_2=A_2+B_3+C_4+D_2-3M.$$

Подставляя сюда значения соответствующих (не до конца вычисленных) частных средних и общего среднего M, получим:

$$W_2=\frac{46+W_2}{8}+\frac{54+W_2}{8}+\frac{44+W_1+W_2}{8}+$$

$$+\frac{38+W_2}{8}-3\left(\frac{180+W_1+W_2}{32}\right)$$

или, после преобразования, окончательно:

$$-W_1+19W_2=188.$$

Эти два уравнения составляют систему:

$$\begin{cases} 19W_1-W_2=28 \\ -W_1+19W_2=188, \end{cases}$$

решая которую, получим искомые значения:

$$W_1 = 2 \quad \text{и} \quad W_2 = 10.$$

Теперь представляется возможным довести до конца вычисление всех частных средних и общего среднего M:

$$A_2 = \frac{46 + W_1}{8} = \frac{46 + 10}{8} = \frac{56}{8} = 7$$

$$A_3 = \frac{22 + W_2}{8} = \frac{22 + 10}{8} = \frac{32}{8} = 4$$

$$B_2 = \frac{46 + W_1}{8} = \frac{46 + 2}{8} = \frac{48}{8} = 6$$

$$B_3 = \frac{54 + W_2}{8} = \frac{54 + 10}{8} = \frac{64}{8} = 8$$

$$C_4 = \frac{44 + W_1 + W_2}{8} = \frac{44 + 2 + 10}{8} = \frac{56}{8} = 7$$

$$D_1 = \frac{22 + W_1}{8} = \frac{22 + 2}{8} = \frac{24}{8} = 3$$

$$D_2 = \frac{38 + W_2}{8} = \frac{38 + 10}{8} = \frac{48}{8} = 6$$

$$M = \frac{180 + W_1 + W_2}{32} = \frac{180 + 2 + 10}{32} = \frac{192}{32} = 6.$$

Ниже приведем расчет сумм квадратов отклонений α, β, γ и δ всех этих частных средних A, B, C и D от общего среднего $M = 6$:

A	α	α^2	B	β	β^2	C	γ	γ^2	D	δ	δ^2
10	+4	16	5	−1	1	7	+1	1	3	−3	9
7	+1	1	6	0	0	6	0	0	6	0	0
4	−2	4	8	+2	4	4	−2	4	6	0	0
3	−3	9	5	−1	1	7	+1	1	9	+3	9
	0	30		0	6		0	6		0	18

$\Sigma\alpha^2 = 30$ $\quad\Sigma\beta^2 = 6$ $\quad\Sigma\gamma^2 = 6$ $\quad\Sigma\delta^2 = 18$

Отсюда получаем суммы:

$$S_a = \frac{N\Sigma\alpha^2}{a} = \frac{32 \cdot 30}{4} = 240$$

$$S_b = \frac{N\Sigma\beta^2}{b} = \frac{32 \cdot 6}{4} = 48$$

$$S_c = \frac{N\Sigma\gamma^2}{c} = \frac{32 \cdot 6}{4} = 48$$

$$S_d = \frac{N\Sigma\delta^2}{d} = \frac{32 \cdot 18}{4} = 144.$$

Для вычисления суммы S_x возвысим предварительно все 30 эмпирических дат в квадрат и прибавим к ним сумму квадратов двух вычисленных теоретических дат $W_1 = 2$ и $W_2 = 10$:

$$\Sigma V^2 = 1562 + W_1^2 + W_2^2 = 1562 + 4 + 100 = 1666,$$

откуда искомая сумма (при $N = 32$ и $M = 6$):

$$S_x = \Sigma V^2 - NM^2 = 1666 - 32 \cdot 6^2 =$$
$$= 1666 - 32 \cdot 36 = 1666 - 1152 = 514.$$

Наконец сумма:

$$S_z = S_x - (S_a + S_b + S_c + S_d) =$$
$$= 514 - (240 + 48 + 48 + 144) = 514 - 480 = 34.$$

Далее определяем число степеней свободы (ν).

Здесь ν_x должно быть уменьшено на 2, так как в числе 32 дат имеются две (W_1 и W_2), искусственно подогнанные нами ко всей совокупности частных средних:

$$\nu_x = (N - 1) - 2 = (32 - 1) - 2 = 31 - 2 = 29$$
$$\nu_a = a - 1 = 4 - 1 = 3 \qquad \nu_c = c - 1 = 4 - 1 = 3$$
$$\nu_b = b - 1 = 4 - 1 = 3 \qquad \nu_d = d - 1 = 4 - 1 = 3$$
$$\nu_z = \nu_x - (\nu_a + \nu_b + \nu_c + \nu_d) = 29 - (3 + 3 + 3 + 3) =$$
$$= 29 - 12 = 17.$$

Ниже дается вычисление мер варьирования:

$$\sigma_a^2 = \frac{S_a}{\nu_a} = \frac{240}{3} = 80 \qquad \sigma_c^2 = \frac{S_c}{\nu_c} = \frac{48}{3} = 16$$
$$\sigma_b^2 = \frac{S_b}{\nu_b} = \frac{48}{3} = 16 \qquad \sigma_d^2 = \frac{S_d}{\nu_d} = \frac{144}{3} = 48$$

$$\sigma_z^2 = \frac{S_z}{\nu_z} = \frac{34}{17} = 2 \qquad \sigma_x^2 = \frac{S_x}{\nu_x} = \frac{514}{29} = 17{,}73.$$

Далее вычисляем критерии Θ (при $\nu_1 = 3$ и $\nu_2 = 17$)

$$\Theta_a = \frac{\sigma_a^2}{\sigma_z^2} = \frac{80}{2} = \underline{\underline{40}} \qquad \Theta_c = \frac{\sigma_c^2}{\sigma_z^2} = \frac{16}{2} = \underline{\underline{8}}$$
$$\Theta_b = \frac{\sigma_b^2}{\sigma_z^2} = \frac{16}{2} = \underline{\underline{8}} \qquad \Theta_d = \frac{\sigma_d^2}{\sigma_z^2} = \frac{48}{2} = \underline{\underline{24}}$$

Для сравнения эффективности отдельных факторов между собою вычисляем соответствующие показатели Θ (при $\nu_1 = 3$ и $\nu_2 = 3$):

$$\Theta_{a-b} = \frac{\sigma_a^2}{\sigma_b^2} = \frac{80}{16} = 5{,}00 \qquad \Theta_{b-c} = \frac{\sigma_b^2}{\sigma_c^2} = \frac{16}{16} = 1{,}00$$
$$\Theta_{a-c} = \frac{\sigma_a^2}{\sigma_c^2} = \frac{80}{16} = 5{,}00 \qquad \Theta_{d-b} = \frac{\sigma_d^2}{\sigma_b^2} = \frac{48}{16} = 3{,}00$$
$$\Theta_{a-d} = \frac{\sigma_a^2}{\sigma_d^2} = \frac{80}{48} = 1{,}67 \qquad \Theta_{d-c} = \frac{\sigma_d^2}{\sigma_c^2} = \frac{48}{16} = 3{,}00$$

(достоверных различий не оказалось).

Квадрат средней ошибки m (в нашем примере одинаковой для всех четырех факторов, так как число вариантов $a = b = c = d = h$) получим по формуле:

$$m^2 = \frac{n\sigma_z^2}{N} = \frac{4 \cdot 2}{32} = 0{,}25,$$

а среднюю ошибку разности любой пары частных средних:

$$\Delta = \sqrt{2m^2} = \sqrt{2 \cdot 0{,}25} = \sqrt{0{,}50} = 0{,}71.$$

По таблице Фишера-Стьюдента для:

$$\nu = \nu_z = 17$$

находим: $t_5 = 2{,}11$ $t_1 = 2{,}90$ $t_{01} = 3{,}97$. Следовательно:

$$l_5 = \Delta t_5 = 0{,}71 \cdot 2{,}11 = 1{,}50$$
$$l_1 = \Delta t_1 = 0{,}71 \cdot 2{,}90 = 2{,}06$$
$$l_{01} = \Delta t_{01} = 0{,}71 \cdot 3{,}97 = 2{,}88.$$

IX

Разобранный пример иллюстрировал схему анализа 4-х факторов, каждый из которых испытывается в 4-х вариантах. Мыслимы очевидно и другие схемы („греко-латинские квадраты“), напр., нижеприводимый комплекс шести факторов A, B, C, D, E и F, заданных в пяти вариантах каждый:

	B_1	B_2	B_3	B_4	B_5
A_1	№ 1 C_1D_1; 6 (E_1) (F_1)	№ 6 C_2D_2; 14 (E_2) (F_2)	№ 11 C_3D_3; W_2 (E_3) (F_3)	№ 16 C_4D_4; 10 (E_4) (F_4)	№ 21 C_5D_5; 8 (E_5) (F_5)
A_2	№ 2 C_2D_3; 5 (E_5) (F_4)	№ 7 C_3D_4; 3 (E_1) (F_5)	№ 12 C_4D_5; 11 (E_2) (F_1)	№ 17 C_5D_1; 4 (E_3) (F_2)	№ 22 C_1D_2; 2 (E_4) (F_3)
A_3	№ 3 C_3D_5; W_1 (E_4) (F_2)	№ 8 C_4D_1; 11 (E_5) (F_3)	№ 13 C_5D_2; 8 (E_1) (F_4)	№ 18 C_1D_3; 0 (E_2) (F_5)	№ 23 C_2D_4; 7 (E_3) (F_1)
A_4	№ 4 C_4D_2; 9 (E_3) (F_5)	№ 9 C_5D_3; 4 (E_4) (F_1)	№ 14 C_1D_4; 9 (E_5) (F_2)	№ 19 C_2D_5; 19 (E_1) (F_3)	№ 24 C_3D_1; 14 (E_2) (F_4)
A_5	№ 5 C_5D_4; 1 (E_2) (F_3)	№ 10 C_1D_5; 13 (E_3) (F_4)	№ 15 C_2D_1; 15 (E_4) (F_5)	№ 20 C_3D_2; 7 (E_5) (F_1)	№ 25 C_4D_3; 4 (E_1) (F_2)

Здесь имеется $5 \times 5 = 25$ клеток, соответствующих отдельным сочетаниям вариантов этих факторов (указанных буквами и цифровыми значками). Нетрудно убедиться в том, что здесь так же, как и в предыдущей таблице (4×4), отдельные варианты каждого фактора входят в комбинацию со всеми вариантами каждого из остальных факторов и притом всюду по одному разу. Таким образом вместо $5 \times 5 \times 5 \times 5 \times 5 \times 5 = 15625$ комбинаций здесь использовано только 25 (делянок, отдельных опытов и т. д.). В клетках за №№ 3 и 11 попрежнему имеются выпадающие даты, заменяемые теоретическими их значениями W_1 и W_2. Схема дана здесь для сочетаний всех шести факторов, но в числовом примере использованы лишь 4 (буквы E и F всюду за ненадобностью поставлены в скобки). В каждой клетке здесь стоит одна дата V (а не две, как в предыдущем примере).

Производя обработку этих данных по описанному выше способу, получим два уравнения:

$$\begin{cases} W_1 = A_3 + B_1 + C_3 + D_5 - 3M \\ W_2 = A_1 + B_3 + C_3 + D_3 - 3M \end{cases}$$

Подставляя сюда значения отдельных средних, будем иметь:

$$W_1 = \frac{26 + W_1}{5} + \frac{21 + W_1}{5} + \frac{24 + W_1 + W_2}{5} + \\ + \frac{51 + W_1}{5} - 3\left(\frac{184 + W_1 + W_2}{25}\right)$$

$$W_2 = \frac{38 + W_1}{5} + \frac{43 + W_2}{5} + \frac{24 + W_1 + W_2}{5} + \\ + \frac{13 + W_2}{5} - 3\left(\frac{184 + W_1 + W_2}{25}\right)$$

или, после преобразования, окончательно:

$$\begin{cases} 4W_1 - W_2 = 29 \\ -W_1 + 4W_2 = 19, \end{cases}$$

откуда найдем: $W_1 = 9$ и $W_2 = 7$.

Далее вычислим все средние $M = 8$:

$A_1 = 9$	$B_1 = 6$	$C_1 = 6$	$D_1 = 10$
$A_2 = 5$	$B_2 = 9$	$C_2 = 12$	$D_2 = 8$
$A_3 = 7$	$B_3 = 10$	$C_3 = 8$	$D_3 = 4$
$A_4 = 11$	$B_4 = 8$	$C_4 = 9$	$D_4 = 6$
$A_5 = 8$	$B_5 = 7$	$C_5 = 5$	$D_5 = 12$

Затем находим суммы:

$$\Sigma\alpha^2 = 20 \qquad \Sigma\gamma^2 = 30$$
$$\Sigma\beta^2 = 10 \qquad \Sigma\delta^2 = 40,$$

а с их помощью определяем и суммы:

$$S_a = 100 \qquad S_c = 150$$
$$S_b = 50 \qquad S_d = 200$$

Так как в нашем примере $\Sigma V^2 = 2000$, то:

$$S_x = 530 \qquad \text{и} \qquad S_z = 30.$$

Число степеней свободы:

$$\nu_x = 24 - 2 = 22 \qquad \nu_a = \nu_b = \nu_c = \nu_d = 4 \qquad \nu_z = 6.$$

Следовательно:

$$\sigma_a^2 = 25 \qquad \sigma_c^2 = 37{,}5 \qquad \sigma_z^2 = 5$$

$$\sigma_b^2 = 12{,}5 \qquad \sigma_d^2 = 50$$

Показатели Θ получаются равными:

$$\Theta_a = 5 \qquad \Theta_c = 7{,}5$$

$$\Theta_b = 2{,}5 \qquad \Theta_d = 10$$

Квадрат средней ошибки и ошибка разности:

$$m^2 = 1 \qquad \Delta = 1{,}41.$$

Значение t ищем для числа степеней свободы:

$$\nu = 6.$$

При этих условиях минимальные разности оказываются равными:

$$l_5 = 3{,}41 \qquad l_1 = 5{,}23 \qquad l_{01} = 9{,}40.$$

Оба разобранные выше образца греко-латинских квадратов (4×4 и 5×5) являлись не только материалом для числовой иллюстрации метода их обработки, но и представляли собой стандарты для планирования аналогичных экспериментов. С этой целью наряду с факторами, действительно использованными в числовом примере, там были указаны и комбинации других факторов, в числовой пример не вошедших. В заключение приведем еще одну схему греко-латинского квадрата (3×3) с участием 4-х факторов, на этот раз без всякой числовой иллюстрации:

	B_1	B_2	B_3
A_1	C_1D_1	C_2D_2	C_3D_3
A_2	C_2D_3	C_3D_1	C_1D_2
A_3	C_3D_2	C_1D_3	C_2D_1

В настоящее время составлены схемы греко-латинских квадратов 8×8 и 9×9, но мы здесь их не приводим, так как надобность в них возникает крайне редко.

X

В применении к полным комплексам признаков (без сокращений по схеме греко-латинского квадрата) удобно использовать особый вариант статистической обработки, известный под наименованием „факториального анализа". Покажем применение этих приемов примером анализа действия двух факторов A и B на результативный признак V при двукратной повторности:

	B_1	B_2	B_3	B_4	
A_1	19 17	29 31	47 49	71 73	336
A_2	11 13	27 33	49 47	55 53	288
A_3	7 5	25 23	31 29	37 35	192
A_4	13 11	11 13	17 19	31 29	144
	96	192	288	384	960

Здесь оба фактора (A и B) заданы в одинаковом числе вариантов ($a=4$ и $b=4$), а каждая из $4 \times 4 = 16$ комбинаций этих вариантов представлена двумя датами (обозначим это число буквой n, у нас $n=2$). Следовательно, общее число всех дат $N = abn = 4 \cdot 4 \cdot 2 = 32$.

Обработаем этот комплекс обычным способом. Здесь сумма всех дат $\Sigma V = 960$, сумма их квадратов $\Sigma V^2 = 38640$ и, следовательно, общее среднее:

$$M = \frac{\Sigma V}{N} = \frac{960}{32} = 30,$$

а общее рассеяние:

$$S_x = \Sigma V^2 - NM^2 = 38640 - 32 \cdot (30)^2 = 38640 - 32 \cdot 900 = 9840.$$

Частные средние окажутся равными:

$$A_1 = \frac{336}{8} = 42 \qquad B_1 = \frac{96}{8} = 12$$

$$A_2 = \frac{288}{8} = 36 \qquad B_2 = \frac{192}{8} = 24$$

$$A_3 = \frac{192}{8} = 24 \qquad B_3 = \frac{288}{8} = 36$$

$$A_4 = \frac{144}{8} = 18 \qquad B_4 = \frac{384}{8} = 48$$

Ниже дан расчет сумм $\Sigma\alpha^2$ и $\Sigma\beta^2$:

A	α	α^2	B	β	β^2
42	+12	144	12	−18	324
36	+ 6	36	24	− 6	36
24	− 6	36	36	+ 6	36
18	−12	144	48	+18	324
		$\Sigma\alpha^2 = 360$			$\Sigma\beta^2 = 720$

Далее вычисляем суммы:

$$S_a = \frac{N\Sigma\alpha^2}{a} = \frac{32 \cdot 360}{4} = 2880$$

$$S_b = \frac{N\Sigma\beta^2}{b} = \frac{32 \cdot 720}{4} = 5760,$$

откуда сумма:

$$S_z = S_x - (S_a + S_b) = 9\,840 - (2\,880 + 5\,760) = 9\,840 - 8\,640 = 1\,200.$$

Число степеней свободы:

$$\nu_x = N - 1 = 32 - 1 = 31$$

$$\nu_a = a - 1 = 4 - 1 = 3$$

$$\nu_b = b - 1 = 4 - 1 = 3$$

$$\nu_z = \nu_x - (\nu_a + \nu_b) = 31 - (3 + 3) = 31 - 6 = 25.$$

Следовательно, квадраты мер варьирования:

$$\sigma_a^2 = \frac{S_a}{\nu_a} = \frac{2\,880}{3} = 960$$

$$\sigma_b^2 = \frac{S_b}{\nu_b} = \frac{5\,760}{3} = 1\,920$$

$$\sigma_z^2 = \frac{S_z}{\nu_z} = \frac{1\,200}{25} = 48.$$

Оценка достоверности этих показателей дает следующие результаты:

$$\Theta_a = \frac{\sigma_a^2}{\sigma_z^2} = \frac{960}{48} = \underline{\underline{20}} \text{ (при } \nu_1 = 3 \text{ и } \nu_2 = 25)$$

$$\Theta_b = \frac{\sigma_b^2}{\sigma_z^2} = \frac{1\,920}{48} = \underline{\underline{40}} \text{ (при } \nu_1 = 3 \text{ и } \nu_2 = 25).$$

Для наших целей здесь достаточно остановиться на этой фазе статистического анализа и запомнить лишь, что квадрат меры случайного варьирования $\sigma_z^2 = 48$.

XI

Проглядывая первоначальные даты V, занесенные в клетки приведенной выше таблицы, мы замечаем, что в верхней ее части (для A_1 и A_2) эти даты возрастают слева направо довольно энергично, в то время как в нижней части (для A_3 и A_4) такое их возрастание наблюдается значительно слабее. Это наводит на мысль, что, объединяя даты V по столбцам (при вычислении частных средних по вариантам фактора B), мы могли смазать эти особенности и не отразить их должным образом в величине полученных частных средних B.

Аналогичное соображение относится и к группировкам наших дат по отдельным вариантам фактора B.

Для проверки этих подозрений рассечем данную таблицу крестообразно на четыре части и в каждой из них объединим все попавшие туда даты V, вычислив для них отдельные средние. Пусть, например, это рассечение нашей таблицы было произведено строго симметрично, и в каждую из четырех частей новой таблицы попало поровну всюду по 8 дат.

	B_1	B_2	B_3	B_4	
A_1	19	29	47	71	336
	17	31	49	73	
A_2	11	27	49	55	288
	13	33	47	53	
A_3	7	25	31	37	192
	5	23	29	35	
A_4	13	11	17	31	144
	11	13	19	29	
	96	192	288	384	960.

После подсчета средних значений данного признака для каждой из четырех клеток этой таблицы получим новую краткую таблицу, состоящую всего лишь из 4-х дат.

	B_{12}	B_{34}	
A_{12}	$\frac{180}{8}=22{,}5$	$\frac{444}{8}=55{,}5$	$\frac{624}{16}=39$
A_{34}	$\frac{108}{8}=13{,}5$	$\frac{228}{8}=28{,}5$	$\frac{336}{16}=21$
	$\frac{288}{16}=18$	$\frac{672}{16}=42$	$\frac{960}{32}=30.$

Здесь на полях, справа и снизу, дано вычисление также и новых частных средних для отдельных столбцов и строчек этой таблицы, а в правом нижнем углу приведен расчет общего среднего M, совпавшего в данном случае с прежним его значением $M = 30$.

Итак, имеем следующую таблицу:

$M = 30$	$B_{12} = 18$	$B_{34} = 42$
$A_{12} = 39$	22,5	55,5
$A_{34} = 21$	13,5	28,5

Вычислим для каждой из 4-х новых дат V их „случайные“ отклонения z путем сравнения наших эмпирических дат V с их теоретическими значениями W, определяемыми по указанной ранее формуле:

$$W = A + B - (f - 1)\ M$$

или, так как у нас число факторов $f = 2$, то окончательно по формуле:

$$W = A + B - M.$$

В применении к левой верхней клетке (где стоит дата $V = 22{,}5$) получим:

$$W = A_{12} + B_{12} - M = 39 + 18 - 30 = 57 - 30 = 27.$$

Для соседней справа клетки (где $V = 55{,}5$)

$$W = A_{12} + B_{34} - M = 39 + 42 - 30 = 81 - 30 = 51.$$

Для левой нижней клетки ($V = 13{,}5$)

$$W = A_{34} + B_{12} - M = 21 + 18 - 30 = 39 - 30 = 9.$$

Для правой нижней клетки ($V = 28{,}5$)

$$W = A_{34} + B_{34} - M = 21 + 42 - 30 = 63 - 30 = 33.$$

Записав все теоретические даты W под соответствующими эмпирическими V и вычитая из V эти значения W, получим „случайные“ отклонения:

$$z = V - W.$$

	B_{12}	B_{34}
A_{12}	$V = 22{,}5$ $W = 27$ $z = -4{,}5$	$V = 55{,}5$ $W = 51$ $z = +4{,}5$
A_{34}	$V = 13{,}5$ $W = 9$ $z = +4{,}5$	$V = 28{,}5$ $W = 33$ $z = -4{,}5$

Здесь во всех четырех клетках получились совершенно одинаковые по своей абсолютной величине отклонения z (всюду 4,5), причем знаки у них оказались чередующимися в определенном порядке: в правой верхней и в левой нижней клетке получился знак $+$, а в левой верхней и в правой нижней знак $-$. Отмеченные особенности будут иметь место в применении ко всякой таблице из 4-х клеток, при условии получения ее при помощи строго симметричного деления первоначальной (большой) таблицы на 4 части. Следовательно, сумма этих отклонений $\Sigma z = 0$.

Возвысив каждое из этих четырех одинаковых отклонений z в квадрат и сложив их, получим сумму:

$$S_z = 4_z^{\,2} = 4\ (4{,}5)^2 = 4 \cdot 20{,}25 = 81.$$

Легко убедиться в том, что для подобного рода четырехклеточных таблиц число степеней свободы ν_z всегда $= 1$.

В самом деле здесь (при $N=4, a=2$ и $b=2$):

$$\nu_x = N-1 = 4-1 = 3$$

$$\nu_a = a-1 = 2-1 = 1$$

$$\nu_b = b-1 = 2-1 = 1$$

$$\nu_z = \nu_x - (\nu_a + \nu_b) = 3-(1+1) = 3-2 = 1.$$

Следовательно, квадрат меры случайного варьирования:

$$\sigma_z^2 = \frac{S_z}{\nu_z} = \frac{81}{1} = 81.$$

Посмотрим, в какой степени этот показатель (σ_z^2) действительно характеризует чисто случайные колебания наших дат, и не произошло ли здесь некоторого закономерного скопления отклонений z в отдельных клетках нашей новой таблицы в результате совместного действия на признак V со стороны обоих факторов A и B одновременно.

Для решения этого вопроса вычислим теоретическое (ожидаемое) значение меры случайных колебаний четырех наших новых дат, полученных в качестве средних из $N=32$ первоначальных дат V. По общим законам случайного варьирования эта мера ожидаемых колебаний таких средних должна оказаться во столько раз меньше меры случайных колебаний наших первоначальных дат V (т. е. $\sigma_z^2 = 48$), во сколько раз количество этих дат (т. е. 4) меньше числа первоначальных дат $N=32$. Так как в нашем случае это новое число дат (4) оказалось в 8 раз меньше прежнего (32), то и ожидаемая мера их случайных колебаний (назовем ее σ_0^2) должна быть также в 8 раз меньше ранее вычисленной ($\sigma_z^2 = 48$). Итак:

$$\sigma_0^2 = \frac{\sigma_z^2}{8} = \frac{48}{8} = 6.$$

Следовательно, если бы распределение отклонений z было чисто случайным, то мера их колеблемости в нашей новой четырехклеточной таблице оказалась бы равной:

$$\sigma_0^2 = 6,$$

а не значению $\sigma_z^2 = 81$, которое фактически получилось у нас при вычислении. Сравнивая это действительное значение $\sigma_z^2 = 81$ с теоретически ожидаемым $\sigma_0^2 = 6$, получим:

$$\Theta = \frac{\sigma_z^2}{\sigma_0^2} = \frac{81}{6} = \underline{\underline{13{,}5}} \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 25).$$

Следовательно, обнаруженное нами скопление отклонений z в отдельных клетках этой таблицы следует признать не случайным, а совместное действие факторов A и B (в данном их подразделении) на результативный признак V считать статистически доказанным с вероятностью, превышающей 0,99.

XII

Аналогично можно было бы исследовать наличие или отсутствие синергизма этих факторов A и B и в любом другом их альтернативном подразделении. Пусть, например, первый фактор (A) мы хотим разбить на две неравночисленные группы, объединив первые три его варианта: A_1, A_2 и A_3 и противопоставив им оставшийся четвертый вариант A_4, а по второму фактору (B) сопоставить его первый вариант B_1 с последними двумя B_3 и B_4, вовсе не принимая при этом во внимание второго варианта B_2.

Такому подразделению будет отвечать следующее объединение наших первоначальных дат V:

	B_1	B_3	B_4
A_1	19 17	47 49	71 73
A_2	11 13	49 47	55 53
A_3	7 5	31 29	37 35
A_4	13 11	17 19	31 29

Здесь они скопились в четырех клетках новой таблицы не в одинаковом количестве: в левой верхней клетке их оказалось 6, в правой верхней — 12, в левой нижней — 2, а в правой нижней — 4. Всего в эту таблицу попало 24 даты (вместо 32), так как целый столбец (B_2) сюда не вошел вовсе. Складывая в каждой клетке все попавшие в нее даты V и деля их сумму на количество этих дат (p), получим 4 средние, являющиеся 4-мя новыми датами этой малой таблицы:

	B_1	B_{34}	
A_{123}	$\frac{72}{6}=12$	$\frac{576}{12}=48$	$\frac{648}{18}=36$
A_4	$\frac{24}{2}=12$	$\frac{96}{4}=24$	$\frac{120}{6}=20$
	$\frac{96}{8}=12$	$\frac{672}{16}=42$	$\frac{768}{24}=32$

Здесь, на полях, дан расчет частных средних по отдельным вариантам факторов A и B, а справа внизу проведено вычисление общего среднего M. Каждое из этих средних получено путем деления соответствующей суммы дат на их общее количество. Например, для среднего A_4 (нижняя строчка) произведено объединение двух внутренних средних:

$$\frac{24}{2}=12 \quad \text{и} \quad \frac{96}{4}=24$$

путем сложения обоих числителей $24+96=120$ и обоих знаменателей $2+4=6$, с последующим делением этой суммы дат (120) на их количество (6):

$$\frac{120}{6}=20.$$

В конечном итоге получается такая таблица:

$M=32$	$B_1=12$	$B_{34}=42$
$A_{123}=36$	12	48
$A_4=20$	12	24

Вычисляя теоретические даты W, получим для верхней левой клетки ($V=12$):

$$W=36+12-32=48-32=16,$$

для правой верхней клетки ($V=48$):

$$W=36+42-32=78-32=46,$$

для левой нижней клетки ($V=12$):

$$W=20+12-32=32-32=0,$$

для правой нижней клетки ($V=24$):

$$W=20+42-32=62-32=30.$$

Вычитая в каждой клетке из V это значение W, получим отклонения z:

	B_1	B_{34}
A_{123}	$V=12$ $W=16$ $z=-4$	$V=48$ $W=46$ $z=+2$
A_4	$V=12$ $W=0$ $z=+12$	$V=24$ $W=30$ $z=-6$

Здесь знаки у z чередуются в прежнем порядке, но абсолютная их величина оказывается различной. Следовательно, сумма этих отклонений:

$$\Sigma z=-4+2+12-6=+4,$$

а не 0, как это получилось в прошлый раз. Это объясняется тем, что наши четыре новые даты (12, 48, 12 и 24) были вычислены не из одинакового числа первоначальных дат V (как раньше), а при разном их количестве p (6, 12, 2 и 4). Помножив отдельные значения z на эти частоты (веса) p, получим взвешенную сумму:

$$\Sigma pz = 6(-4) + 12(+2) + 2(+12) + 4(-6) = -24 +$$
$$+ 24 + 24 - 24 = 0,$$

которая всегда окажется равной 0 (нулю).

Ранее (при полном равенстве всех z по их абсолютной величине) мы вычисляли сумму S_z по формуле:

$$S_z = 4z^2.$$

Теперь вместо z^2 следует получить взвешенное среднее их квадратов путем деления взвешенной суммы квадратов всех четырех z на сумму этих весов (p). У нас взвешенная сумма квадратов:

$$\Sigma pz^2 = 6(-4)^2 + 12(+2)^2 + 2(+12)^2 + 4(-6)^2 =$$
$$= 6 \cdot 16 + 12 \cdot 4 + 2 \cdot 144 + 4 \cdot 36 =$$
$$= 96 + 48 + 288 + 144 = 576.$$

Разделив эту сумму (576) на сумму весов (частот) p:

$$\Sigma p = 6 + 12 + 2 + 4 = 24,$$

получим взвешенное среднее:

$$\frac{\Sigma pz^2}{\Sigma p} = \frac{576}{24} = 24.$$

Следовательно, искомая сумма:

$$S_z = 4\left[\frac{\Sigma pz^2}{\Sigma p}\right] = 4 \cdot 24 = 96.$$

Так как число степеней свободы попрежнему $\nu_z = 1$, то:

$$\sigma_z^2 = \frac{S_z}{\nu_z} = \frac{96}{1} = 96.$$

Теоретическое (ожидаемое) значение этой меры (σ_0^2) найдем из тех же соображений, как и в применении к предыдущему случаю. Наши 4 новые даты получены путем объединения в 4 клетках 24 прежних дат, следовательно, количество их уменьшилось:

$$\text{в } \frac{24}{4} = 6 \text{ раз.}$$

Отсюда и мера случайных колебаний ($\sigma_z^2 = 48$) должна была бы уменьшиться в 6 раз, т. е.:

$$\sigma_0^2 = \frac{48}{6} = 8.$$

Итак, если бы все первоначальные отклонения z были действительно случайными, то после объединения в 4-х клетках нашей новой таблицы мера их колеблемости оказалась бы $\sigma_0^2 = 8$. Фактически же мы обнаружили в этой таблице значительно более сильное колебание ($\sigma_z^2 = 96$), в силу чего эту гипотезу случайного их происхождения приходится отвергнуть, ибо:

$$\Theta = \frac{\sigma_z^2}{\sigma_0^2} = \frac{96}{8} = 12 \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 25).$$

Итак, наши два фактора A и B проявляют синергизм своего действия на признак V также и в данном своем неравномерном разделении, порождающем указанное выше несимметричное сечение первоначальной таблицы.

XIII

Путем аналогичных соображений и приемов можно оценить достоверность различия между двумя объединенными средними в пределах одного фактора. Пусть, например, в том же комплексе признаков мы пожелали бы сопоставить средний результат действия первых двух вариантов фактора A (A_1 и A_2, объединяемых символом A_{12}) со средним результатом остальных двух его вариантов (A_3 и A_4 или короче: A_{34}). Рассекая нашу первоначальную таблицу на две равновеликие части, получим две группы объединенных дат V по 16 дат в каждой. Вот это сечение:

		B_1	B_2	B_3	B_4		
A_{12}	A_1	19 17	29 31	47 49	71 73	336	624
	A_2	11 13	27 33	49 47	55 53	288	
A_{34}	A_3	7 5	25 23	31 29	37 35	192	336
	A_4	13 11	11 13	17 19	31 29	144	
		96	192	288	384	960.	

Вычисляя средние для каждой из этих двух групп, получим:

A_{12}	$\frac{624}{16} = 39$	$\frac{960}{32} = 30 = M$
A_{34}	$\frac{336}{16} = 21$	

Здесь мы имеем всего лишь 2 новые даты V (39 и 21), варьирующие около своего общего среднего $M = 30$. В этом варьировании нет возможности различить „случайный“ компанент, так как за отсутствием учета других факторов (кроме A) не представляется возможным использовать формулу для вычисления теоретических дат W. Поэтому фактическое варьирование двух наших дат V (39 и 21) около их среднего $M = 30$ и будет здесь служить предметом сравнения с ранее выясненной мерой случайного варьирования $\sigma_z^2 = 48$.

У нас дата 39 отклоняется от 30 на $+9$, а дата 21 на -9. Квадраты этих отклонений $= 81$, а сумма их $2 \cdot 81 = 162$. Так как при варьировании двух дат около своего среднего имеется лишь 1 степень свободы, то эту сумму мы должны разделить на 1. Таким образом фактическое варьирование этих дат (неотделимое здесь от „случайного") характеризуется показателем:

$$\sigma_z^2 = 162.$$

Так как число наших новых дат (2) по сравнению с общим числом первоначальных дат (32) оказалось в 16 раз меньше, то во столько же раз должна была бы получиться меньше своего прежнего значения (48) и мера их случайных колебаний:

$$\sigma_0^2 = \frac{48}{16} = 3.$$

Итак, если бы между $A_{12} = 39$ и $A_{34} = 21$ не было бы никакого существенного различия, то мера их колеблемости около $M = 30$ оказалась бы $= 3$, а не 162, как у нас получилось в действительности. Отношение:

$$\Theta = \frac{\sigma_z^2}{\sigma_0^2} = \frac{162}{3} = \underline{\underline{54}} \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 25)$$

и служит характеристикой существенности этого различия и его статистическим доказательством.

Аналогично найдем величину этого показателя Θ и для симметричного сечения первоначальной таблицы по признаку B:

	B_1	B_2	B_3	B_4	
A_1	19 17	29 31	47 49	71 73	336
A_2	11 13	27 33	49 47	55 53	288
A_3	7 5	25 23	31 29	37 35	192
A_4	13 11	11 13	17 19	31 29	144
	96	192	288	384	960

Вычисляя средние, получим:

B_{12}	B_{34}	
$\frac{288}{16} = 18$	$\frac{672}{16} = 42$	$M = \frac{960}{32} = 30$

Отклонения этих дат (18 и 42) от $M=30$ равны здесь — 12 и $+12$, квадраты $=144$, а сумма квадратов $2 \cdot 144=288$. Так как число степеней свободы здесь также $=1$, то мера варьирования:

$$\sigma_z^2=\frac{288}{1}=288.$$

Ожидаемая же величина этой меры очевидно попрежнему:

$$\sigma_0^2=3.$$

Следовательно:

$$\Theta=\frac{\sigma_z^2}{\sigma_0^2}=\frac{288}{3}=\underline{\underline{96}} \text{ (при } \nu_1=1 \text{ и } \nu_2=25).$$

Заметим, что описанный здесь способ оценки достоверности различия между двумя средними A_{12} и A_{34} или B_{12} и B_{34} ничего принципиально нового в метод этой оценки не вносит и представляет собой лишь несколько иной технический ее прием. Покажем, что к тому же самому выводу (о достоверности, например, различия между $A_{12}=39$ и $B_{34}=21$) мы можем притти и обычным путем. Обозначим наши средние (39 и 21), как всегда, буквами M_1 и M_2, их средние ошибки — m_1 и m_2, меру варьирования — σ, а число наблюдений соответственно n_1 и n_2. В нашем примере, следовательно:

$$M_1=39 \qquad n_1=16 \qquad \sigma^2=48.$$
$$M_2=21 \qquad n_2=16$$

Таким образом средние ошибки (их квадраты):

$$m_1^2=\frac{48}{16}=3.$$

$$m_2^2=\frac{48}{16}=3.$$

Ошибка разности этих средних:

$$\sqrt{m_1^2+m_2^2}=\sqrt{3+3}=\sqrt{6}.$$

Сравнивая данную разность наших средних:

$$M_1-M_2=39-21=+18$$

с этой ее средней ошибкой, получим отношение:

$$t=\frac{M_1-M_2}{\sqrt{m_1^2+m_2^2}}=\frac{39-21}{\sqrt{3+3}}=\frac{18}{\sqrt{6}},$$

которое удобнее вычислить, возвысив его предварительно в квадрат:

$$t^2=\left(\frac{18}{\sqrt{6}}\right)^2=\frac{18^2}{6}=\frac{394}{6}=54.$$

Это значение t^2 здесь полностью совпало с ранее полученным значением критерия Θ, который у нас также оказался $=54$:

$$t^2=\Theta=54.$$

Такое совпадение будет иметь место всегда. В соответствии с этим в таблице Фишера-Стьюдента числа столбца Θ при $\nu_1 = 1$ представляют собой квадраты чисел столбца t (при том же числе степеней свободы ν_2).

XIV

При несимметричном сечении первоначальной таблицы ход рассуждений в основном остается без изменений. Пусть требуется сопоставить первые три варианта фактора A (т. е. A_1, A_2 и A_3 с четвертым (A_4). Вот отвечающее этому случаю сечение:

	B_1	B_2	B_3	B_4	
A_1	19 17	29 31	47 49	71 73	336
A_2	11 13	27 33	49 47	55 53	288
A_3	7 5	25 23	31 29	37 35	192
A_4	13 11	11 13	17 19	31 29	144
	96	192	288	384	960

Здесь в верхнюю часть таблицы попали 24 даты, а в нижнюю только 8. Вычисляя средние, получим:

A_{123}	$\frac{816}{24} = 34$
A_4	$\frac{144}{8} = 18$

$$M = \frac{960}{32} = 30.$$

Эти новые даты (34 и 18) отклоняются от $M = 30$ на $+4$ и -12, причем сумма этих отклонений $(4 - 12 = -8)$ не $= 0$. Вычисляя взвешенную их сумму, получим:

$$24(+4) + 8(-12) = +96 - 96 = 0.$$

Квадраты этих отклонений ($+4$ и -12) соответственно равны 16 и 144. Их взвешенное среднее:

$$\frac{24 \cdot 16 + 8 \cdot 144}{24 + 8} = \frac{384 + 1152}{32} = \frac{1536}{32} = 48,$$

откуда искомая сумма квадратов (удвоенная величина среднего квадрата):

$$2 \cdot 48 = 96.$$

Так как число степеней свободы $= 1$, то:

$$\sigma_z^2 = 96.$$

Теоретически ожидаемое значение меры случайного варьирования (σ_0^2) будет во столько же раз меньше ранее найденного показателя (48), во сколько раз число новых дат (2) меньше первоначального их числа (32), т. е. в 16 раз.

Итак, имеем попрежнему:

$$\sigma_0^2 = \frac{48}{16} = 3,$$

откуда искомый показатель:

$$\Theta = \frac{\sigma_z^2}{\sigma_0^2} = \frac{96}{3} = \underline{\underline{32}} \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 25).$$

При желании сопоставить вариант B_1 с вариантами B_3 и B_4 (минуя B_2) получим сечение:

	B_1	B_2	B_3	B_4
A_1	19 17	29 31	47 49	71 73
A_2	11 13	27 33	49 47	55 53
A_3	7 5	25 23	31 29	37 35
A_4	13 11	11 13	17 19	31 29
	96	192	288	384

Здесь в левой части скопилось 8 дат, а в правой 16 (всего 24). Вот расчет средних:

B_1	B_{34}	
$\frac{96}{8} = 12$	$\frac{672}{16} = 42$	$\frac{768}{24} = 32 = M.$

Отклонения этих дат (12 и 42) от среднего $M = 32$ соответственно равны -20 и $+10$, а квадраты их $= 400$ и 100. Взвешенные суммы:

$$8(-20) + 16(+10) = -160 + 160 = 0,$$

а взвешенное среднее квадратов:

$$\frac{8 \cdot 400 + 16 \cdot 100}{8 + 16} = \frac{3200 + 1600}{24} = \frac{4800}{24} = 200.$$

Следовательно, сумма квадратов (удвоенная величина среднего квадрата):

$$2 \cdot 200 = 400.$$

При числе степеней свободы $= 1$ получим:

$$\sigma_z^2 = \frac{400}{1} = 400.$$

Теоретически ожидаемая мера случайных колебаний (σ_0^2) должна быть во столько раз меньше ранее полученной (48), во сколько раз наше число дат (2) меньше первоначального (24), т. е. в 12 раз. Итак:

$$\sigma_0^2 = \frac{48}{12} = 4,$$

откуда искомый показатель:

$$\Theta = \frac{\sigma_z^2}{\sigma_0^2} = \frac{400}{4} = \underline{\underline{100}} \text{ (при } \nu_1 = 1 \text{ и } \nu_2 = 25).$$

XV

Аналогичными приемами можно было бы оценить достоверность различий между любыми группировками отдельных вариантов каждого фактора и исследовать эффективность совместного влияния на V любой комбинации вариантов двух и даже нескольких факторов одновременно. Однако столь громоздкий способ этого анализа делает его практически мало удобным.

Покажем другие приемы такого факториального анализа, всегда приводящие к тому же самому конечному результату, но значительно более коротким (хотя и менее наглядным) путем.

Вспомним, что в нашей первоначальной таблице каждая из $4 \times 4 = 16$ комбинаций факторов A и B была представлена $n = 2$ датами V (всего их было $N = 32$ даты). Объединим прежде всего эти повторения и составим новую таблицу, в 16 клетках которой запишем суммы (S) этих первоначальных дат:

	B_1	B_2	B_3	B_4
A_1	36	60	96	144
A_2	24	60	96	108
A_3	12	48	60	72
A_4	24	24	86	60

Теперь у нас имеется столько же дат (вернее их сумм S), сколько задано сочетаний вариантов фактора A с вариантами фактора B, т. е. только 16 чисел.

Расположим эти суммы (S) в один вертикальный ряд, а слева укажем те сочетания вариантов A и B наших факторов, которые относятся к каждой данной сумме. Справа же оставим место для последующих записей. Вот этот новый вид нашей таблицы (см. стр. 53 сверху).

Теперь зададимся теми комбинациями вариантов отдельных факторов, которые мы хотим оценить и проанализировать. Пусть, по-прежнему, требуется сравнить A_{12} с A_{34} и B_{12} с B_{34} и испытать эффективность совместного действия этих двух факторов (A и B) в данном их симметричном подразделении. В соответствии с этим заданием проставим в первых трех вертикальных столбцах этой таблицы особые числа (обозначим их буквами e) в следующем порядке:

A_1	B_1	36						
	B_2	60						
	B_3	96						
	B_4	144						
A_2	B_1	24						
	B_2	60						
	B_3	96						
	B_4	108						
A_3	B_1	12						
	B_2	48						
	B_3	60						
	B_4	72						
A_4	B_1	24						
	B_2	24						
	B_3	36						
	B_4	60						

В первом столбце против всех комбинаций, в которые входят варианты A_1 и A_2, следует записать значение чисел $e = +1$, а для комбинаций с участием A_3 и A_4 — записать числа $e = -1$. В нашей таблице комбинациями с участием A_1 и A_2 заняты 8 верхних строчек, а комбинациями, в которые входят A_3 и A_4, остальные 8 нижерасположенных строк. В соответствии с этим наш первый столбец будет заполнен в верхней своей половине числами $e = +1$, а в нижней $e = -1$ (см. таблицу).

Второй столбец мы предназначим для указания способа подразделения на две группы ($+1$ и -1) второго признака B. Для этого в тех клетках, которые относятся к вариантам B_1 и B_2, мы проставим числа $e = +1$, а в клетках, относящихся к B_3 и B_4, значение $e = -1$. У нас эти числа ($+1$ и -1) будут чередоваться попарно (два раза $+1$ и два раза -1, считая сверху вниз).

Для сочетаний A и B следует составить отдельный (третий) столбец, записав в него новые числа e, полученные путем попарного перемножения соответствующих чисел e в двух предыдущих столбцах (с соблюдением правила знаков).

В заголовках первых двух столбцов мы поставим следующие символические обозначения способа записи в них чисел e

$+A_1$	$+B_1$
$+A_2$	$+B_2$
$-A_3$	$-B_3$
$-A_4$	$-B_4$

обозначив их сверху в скобках для кратости (A_I) и (B_I), а в заголовке последнего столбца запишем эти два символические обозначения (друг под другом) в знак того, что в этом столбце объединены два предыдущие (путем перемножения чисел e).

Следующие столбцы этой таблицы мы заполним в соответствии с планом дальнейшего факториального анализа.

		S	(A_I) $+A_1$ $+A_2$ $-A_3$ $-A_4$	(B_I) $+B_1$ $+B_2$ $-B_3$ $-B_4$	(A_I) (B_I)			
A_1	B_1	36	+1	+1	+1			
	B_2	60	+1	+1	+1			
	B_3	96	+1	−1	−1			
	B_4	144	+1	−1	−1			
A_2	B_1	24	+1	+1	+1			
	B_2	60	+1	+1	+1			
	B_3	96	+1	−1	−1			
	B_4	108	+1	−1	−1			
A_3	B_1	12	−1	+1	−1			
	B_2	48	−1	+1	−1			
	B_3	60	−1	−1	+1			
	B_4	72	−1	−1	+1			
A_4	B_1	24	−1	+1	−1			
	B_2	24	−1	+1	−1			
	B_3	36	−1	−1	+1			
	B_4	60	−1	−1	+1			

Пусть попрежнему требуется подразделить фактор A на две несимметричные группы, сбъединив варианты A_1, A_2 и A_3 для сопоставления их с четвертым вариантом A_4. Чтобы уравнять эти две неравночисленные группы вариантов (для их сопоставления друг с другом) мы должны будем вторую часть этого подразделения (т. е. A_4) принять во внимание в троекратной степени (так как в первую часть A_{123} входят три варианта). Следовательно, при сопоставлении вариантов A_1, A_2 и A_3 с вариантом A_4 следует первые три взять в ординаре ($e = 1$), а последний втройне ($e = 3$) и приписать этим группам разные знаки (первым трем $+$, а последнему — или наоборот). Обозначив такое подразделение фактора A условным символом A_{II} (в отличие от предыдущего равномерного подразделения, ранее обозначенного A_I), получим следующее обозначение, являющееся заголовком нового (четвертого) столбца:

$$\boxed{\begin{array}{l} (A_{II}) \\ +A_1 \\ +A_2 \\ +A_3 \\ -3A_4 \end{array}}$$

В соответствии с указанием этого заголовка, в данном столбце следует для всех комбинаций, в которые входят варианты A_1, A_2 и A_3, поставить числа $e = +1$, а в тех клетках, которые относятся к варианту A_4, записать значения $e = -3$ (см. таблицу)

Для неравномерного подразделения второго фактора B на B_1 и $B_3 + B_4$ (без B_2) получим очевидно такую символическую запись:

$$\boxed{\begin{matrix}(B_{II})\\ +2B_1\\ -B_2\\ -B_4\end{matrix}}$$

В соответствии с этим заголовком пятого столбца в клетках его следует поставить значение $e = +2$ для комбинаций с участием B_1, $e = 0$ для комбинаций, в которые входит неучитываемый здесь вариант B_2 и по $e = -1$ в клетках для B_3 и B_4. Для исследования совместного действия этих факторов в данном их неравномерном подразделении следует заготовить шестой столбец с заголовком:

$$\boxed{\begin{matrix}(A_{II})\\ (B_{II})\end{matrix}}$$

В этом столбце нужно записать такие значения чисел e, которые получатся в результате попарного перемножения чисел e в двух предыдущих столбцах A_{II} и B_{II}.

		S	(A_I) $+A_1$ $+A_2$ $-A_3$ $-A_4$	(B_I) $+B_1$ $+B_2$ $-B_3$ $-B_4$	(A_1) (B_I)	(A_{II}) $+A_1$ $+A_2$ $+A_3$ $-3A_4$	(B_{II}) $+2B_1$ $-B_3$ $-B_4$	(A_{II}) (B_{II})
A_1	B_1	36	+ 1	+ 1	+ 1	+ 1	+ 2	+ 2
	B_2	60	+ 1	+ 1	+ 1	+ 1	0	0
	B_3	96	+ 1	− 1	− 1	+ 1	− 1	− 1
	B_4	144	+ 1	− 1	− 1	+ 1	− 1	− 1
A_2	B_1	24	+ 1	+ 1	+ 1	+ 1	+ 2	+ 2
	B_2	60	+ 1	+ 1	+ 1	+ 1	0	0
	B_3	96	+ 1	− 1	− 1	+ 1	− 1	− 1
	B_4	108	+ 1	− 1	− 1	+ 1	− 1	− 1
A_3	B_1	12	− 1	+ 1	− 1	+ 1	+ 2	+ 2
	B_2	48	− 1	+ 1	− 1	+ 1	0	0
	B_3	60	− 1	− 1	+ 1	+ 1	− 1	− 1
	B_4	72	− 1	− 1	+ 1	+ 1	− 1	− 1
A_4	B_1	24	− 1	+ 1	− 1	− 3	+ 2	− 6
	B_2	24	− 1	+ 1	− 1	− 3	0	0
	B_3	36	− 1	− 1	+ 1	− 3	− 1	+ 3
	B_4	60	− 1	− 1	+ 1	− 3	− 1	+ 3
Σe^2			16	16	16	48	24	72
ΣeS			+ 288	− 384	− 144	+ 384	− 480	− 288
$J = \frac{(\Sigma eS)^2}{\Sigma e^2}$			5184	9216	1296	3072	9600	1152

После заполнения всех клеток этой таблицы числами e, следует проверить их, убедившись в том, что в каждом столбце сумма:

$$\Sigma e = 0.$$

Далее в каждом столбце подсчитывается сумма квадратов чисел e (Σe^2), и результаты этого подсчета заносятся в первую итоговую строчку таблицы, под числами соответствующих столбцов.

Вторая итоговая строчка заполняется суммами произведений чисел e на суммы S (ΣeS). Так, в частности, умножая записанные слева суммы S (36, 60, 96 и т. д.) на числа e, стоящие в первом (крайнем слева) столбце, озаглавленном сверху символом A_1 и учитывая при этом знаки чисел e ($+$ или $-$), мы получим часть этих произведений (S на e) со знаком $+$, а часть со знаком $-$ (у нас здесь $+624$ и -336).

Окончательную (алгебраическую) сумму этих итогов:

$$624 - 336 = +288$$

мы и запишем внизу этого столбца во второй итоговой строчке, озаглавленной слева обозначением: ΣeS.

Аналогично вычислены и остальные суммы произведений S на e (-384, -144, $+384$ и т. д.), записанные в соответствующих клетках этой второй итоговой строчки.

Последняя итоговая строчка заполняется путем возвышения в квадрат чисел второй (предпоследней) строчки (ΣeS), с последующим делением этих квадратов на числа, стоящие в первой итоговой строчке (Σe^2). Так, в частности, возвышая в квадрат сумму $\Sigma eS = 288$, получим $(\Sigma eS)^2 = (288)^2 = 82944$, а деля ее на $\Sigma e^2 = 16$, получим окончательно:

$$\frac{(\Sigma eS)^2}{\Sigma e^2} = \frac{82944}{16} = 5184.$$

Аналогично получены и остальные числа, записанные в клетках этой последней итоговой строчки (обозначим их буквами J).

С помощью этих чисел J мы легко можем получить те показатели Θ, которые ранее были вычислены в применении ко всем этим шести сопоставлениям отдельных вариантов наших факторов A и B и их совместным сочетаниям. Вспомним, что для этих шести случаев мы получили такие значения показателей:

$$\Theta: \; 54 \quad 96 \quad 13{,}5 \quad 32 \quad 100 \quad 12.$$

Вспомним, далее, что из основной таблицы мы вычислили меру случайных колебаний всех $N = 32$ первоначальных дат:

$$\sigma_z^2 = 48,$$

и что каждая из 16 комбинаций четырех вариантов фактора A с четырьмя вариантами фактора B была представлена двумя датами V, т. е. что повторность в этом опыте:

$$n = 2.$$

Так вот, если найденные только что числа J разделить на произведение:

$$n\sigma_z^2 = 2 \cdot 48 = 96,$$

то мы и получим значения наших показателей Θ.

В самом деле:

$$\frac{5184}{96} = 54 \qquad \frac{3072}{96} = 32$$

$$\frac{9216}{96} = 96 \qquad \frac{9600}{96} = 100$$

$$\frac{1296}{96} = 13{,}5 \qquad \frac{1152}{96} = 12$$

Далее мы оценивали достоверность всех этих показателей Θ по таблице Фишера-Стьюдента при числе степеней свободы $\nu_1 = 1$ (всегда) и $\nu_2 = 25$ (в нашем случае). В соответствующей клетке этой таблицы находим следующие три значения Θ:

$\Theta_5 = 4{,}24$, соответствующее вероятности $P_5 = 0{,}95$

$\Theta_1 = 7{,}77$ „ „ $P_1 = 0{,}99$

$\Theta_{01} = 13{,}88$ „ „ $P_{01} = 0{,}999$

Сравнивая 6 фактически полученных нами значений Θ с этими тремя их ориентировочными значениями, мы видим, что некоторые из них превышают даже $\Theta_{01} = 13{,}88$ (3 подчеркивания), в то время как другие (13,5 и 12) оказываются больше лишь второго значения $\Theta_1 = 7{,}77$ (2 подчеркивания)

$$\underline{\underline{\underline{54}}} \quad \underline{\underline{\underline{96}}} \quad \underline{\underline{13{,}5}} \quad \underline{\underline{\underline{32}}} \quad \underline{\underline{\underline{100}}} \quad \underline{\underline{12}}.$$

Нетрудно убедиться в том, что для выполнения этой оценки нет необходимости в действительности вычислять все показатели Θ путем деления чисел J на $n\sigma_z^2$ (у нас $= 96$). Можно, наоборот, помножить взятые из таблицы Фишера-Стьюдента три значения Θ (4,24, 7,77 и 13,88) на то же самое число $n\sigma_z^2 = 2{,}48 = 96$ и полученные произведения:

$$J_5 = \Theta_5\,(n\sigma_z^2) = 4{,}24 \cdot 96 = 407{,}04$$

$$J_1 = \Theta_1\,(n\sigma_z^2) = 7{,}77 \cdot 96 = 745{,}92$$

$$J_{01} = \Theta_{01}\,(n\sigma_z^2) = 13{,}88 \cdot 96 = 1332{,}48$$

сравнивать со всеми фактически полученными числами J, записанными в последней строчке нашей таблицы:

$$J{:}\ \underline{\underline{\underline{5184}}} \quad \underline{\underline{\underline{9216}}} \quad \underline{\underline{1296}} \quad \underline{\underline{\underline{3072}}} \quad \underline{\underline{\underline{9600}}} \quad \underline{\underline{1152}}$$

Те числа J, которые превышают $J_{01} = 1332{,}48$, мы подчеркиваем 3 раза, а те их значения (1296 и 1152), которые оказываются больше $J_1 = 745{,}92$, подчеркиваем 2 раза и т. д.

XVII

Приведем пример полного факториального анализа комплекса признаков, указав попутно также и способы контроля всех промежуточных вычислений. Этим же примером мы воспользуемся и для иллюстрации метода преобразования первоначальных дат, сглаживающего их колебания, связанные с их величиной, а также еще раз продемонстрируем прием восстановления выпадающих дат в наиболее общем его виде. Пусть имеем следующую таблицу:

	B_1		B_2	
	C_1	C_2	C_1	C_2
A_0	26,54 30,86 21,54	34,31 — 33,14	45,74 35,50 40,46	66,74 73,46 60,34
A_1	26,54 29,75 23,51	4,34 3,11 5,75	26,54 22,54 30,86	0,50 0,94 0,14
A_2	0,50 0,31 0,71	7,34 9,11 5,75	11,75 10,39 11,06	2,74 1,46 2,06

Здесь мы имеем $f=3$ фактора (A, B и C) с числом вариантов $a=3$, $b=2$, $c=2$, при числе повторений $n=3$, всего, следовательно, $N=abcn=3\cdot 2\cdot 2\cdot 3=36$ дат, из которых 35 дат нам известны, а 1 — выпала (в таблице поставлена черта). Прежде чем заняться восстановлением этой выпавшей даты, выполним некоторое их преобразование, для устранения излишних колебаний этих дат, связанных с их величиной.

Дело в том, что рассеяние дат, измеряемое суммой квадратов их отклонений от среднего и являющееся предметом нашего анализа, может вызываться двоякого рода причинами. С одной стороны, варьирование этих дат обусловлено влиянием отдельных изучаемых в опыте факторов, а с другой оно вызывается различиями в самой их величине (известно, что большие даты обычно варьируют сильнее малых). Поскольку второй компонент этого варьирования для нас неинтересен и даже опасен в качестве момента, искажающего колебания основные, связанные с влиянием отдельных факторов, представляется целесообразным варьирования этого рода по возможности уменьшить. Такое сглаживание резких колебаний, связанных с величиной самих дат, достигается применением поправки Бартлетта, осуществляемой при помощи формулы:

$$V=\sqrt{V_1+0{,}5},$$

где V_1 — неисправленная (непосредственно из опыта полученная) дата, а V — искомая дата, исправленная по Бартлетту.

Так, в нашем примере в крайней слева верхней клетке стоят три даты, первая из которых:

$$V_1 = 26{,}54.$$

Для замены ее исправленной датой V производим вычисление по указанной выше формуле:

$$V = \sqrt{26{,}54 + 0{,}50} = \sqrt{27{,}04} = 5{,}2.$$

Выполнив аналогичные вычисления со всеми остальными датами V_1 нашей первоначальной таблицы и заменив их новыми датами V, получим следующую окончательную таблицу:

	B_1		B_2		
	C_1	C_2	C_1	C_2	
A_0	5,2 5,6 4,8	5,9 W 5,8	6,8 6,0 6,4	8,2 8,6 7,8	71,1 + W
A_1	5,2 5,5 4,9	2,2 1,9 2,5	5,2 4,8 5,6	1,0 1,2 0,8	40,8
A_2	1,0 0,9 1,1	2,8 3,1 2,5	3,5 3,3 3,4	1,8 1,4 1,6	26,4
	34,2	26,7 + W	45,0	32,4	138,3 + W

Произведя эту замену, займемся восстановлением выпавшей даты W. Так как дата эта возникла здесь под действием факторов A_0, B_1 и C_2, то теоретическая ее величина:

$$W = A_0 + B_1 + C_2 - (f-1)\,M$$

или, так как у нас число факторов $f = 3$, то:

$$W = A_0 + B_1 + C_2 - 2M.$$

Подсчитав суммы дат V по столбцам и строчкам, а также и общую сумму всех дат (138,3), найдем выражения для всех нужных нам частных средних A_0, B_1 и C_2 и для общего среднего M:

$$A_0 = \frac{71{,}1 + W}{12} \quad C_2 = \frac{26{,}7 + 32{,}4 + W}{18} = \frac{59{,}1 + W}{18}$$

$$B_1 = \frac{34{,}2 + 26{,}7 + W}{18} = \frac{60{,}9 + W}{18} \quad M = \frac{138{,}3 + W}{36}$$

Подставляя эти выражения в наше последнее уравнение, получим:

$$W = \frac{71{,}1 + W}{12} + \frac{60{,}9 + W}{18} + \frac{59{,}1 + W}{18} - 2\left(\frac{138{,}3 + W}{36}\right)$$

Решая это уравнение относительно неизвестного W, в конечном итоге получим:

$$W = 5{,}7.$$

Отсюда легко получаем значения всех частных средних и общего среднего M:

$A_0 = 6{,}4$
$A_1 = 3{,}4$ $B_1 = 3{,}7$ $C_1 = 4{,}4$ $M = 4.$
$A_2 = 2{,}2$ $B_2 = 4{,}3$ $C_2 = 3{,}6$

Для вычисления суммы квадратов всех дат V воспользуемся суммой первоначальных (не исправленных еще по Бартлетту) дат V_1. В нашем примере:

$$\Sigma V_1 = 707{,}33.$$

Чтобы получить сумму квадратов всех наших новых (исправленных) дат V (т. е. ΣV^2), следует воспользоваться формулой:

$$\Sigma V^2 = \Sigma V_1^2 + \frac{N}{2} + \Sigma (W^2 - 0{,}5).$$

В нашем примере была восстановлена лишь одна дата ($W = 5{,}7$). Следовательно, искомая сумма:

$$\Sigma V^2 = 707{,}33 + \frac{36}{2} + \left(5{,}7^2 - 0{,}5\right) = 707{,}33 + 18 + 32{,}49 - 0{,}5 =$$
$$= 757{,}32.$$

После этого определяем сумму:

$$S_x = \Sigma V^2 - NM^2 = 757{,}32 - 36 \cdot (4)^2 = 181{,}32.$$

Ниже приведен расчет отклонений α, β и γ:

A	α	α^2
6,4	+ 2,4	5,76
3,4	− 0,6	0,36
2,2	− 1,8	3,24
		$\Sigma\alpha^2 = 9{,}36$

B	β	β^2
3,7	− 0,3	0,09
4,3	+ 0,3	0,09
		$\Sigma\beta^2 = 0{,}18$

C	γ	γ^2
4,4	+ 0,4	0,16
3,6	− 0,4	0,16
		$\Sigma\beta^2 = 0{,}32$

Далее вычисляем суммы:

$$S_a = \frac{36 \cdot 9{,}36}{3} = 112{,}32$$

$$S_b = \frac{36 \cdot 0{,}18}{2} = 3{,}24$$

$$S_e = \frac{36 \cdot 0{,}32}{2} = 5{,}76$$

Следовательно, сумма:

$$S_z = S_x - (S_a + S_b + S_c) = 181{,}32 - (112{,}32 + 3{,}24 + 5\ 76) =$$
$$= 181{,}32 - 121{,}32 = 60.$$

Числа степеней свободы:

$$\nu_x = (N - 1) - 1 - = 36 - 2 = 34$$

(так как одна дата здесь искусственно восстановлена):

$$\nu_a = a - 1 = 3 - 1 = 2$$
$$\nu_b = b - 1 = 2 - 1 = 1$$
$$\nu_c = c = 1 = 2 - 1 = 1$$
$$\nu_z = \nu_x - (\nu_a + \nu_b + \nu_c) = 34 - (2 + 1 + 1) = 34 - 4 = 30.$$

Следовательно, квадрат меры случайного варьирования:

$$\sigma^2_z = \frac{S_z}{\nu_z} = \frac{60}{30} = 2.$$

Дальнейший анализ этой таблицы будем продолжать по факториальной схеме.

XVIII

Объединив в нашей первоначальной таблице отдельные даты V, соответствующие $n = 3$ повторениям одной и той же комбинации факторов A, B и C, получим следующую сокращенную таблицу значений сумм S:

	B_1		B_2	
	C_1	C_2	C_1	C_2
A_0	15,6	17,4	19,2	24,6
A_1	15,6	6,6	15,6	3,0
A_2	3,0	8,4	10,2	4,8

Здесь фактор A, распадающийся на $a = 3$ варианта (A_0, A_1 и A_2) можно разбить на две части двумя способами, например, сопоставить A_0 с двумя другими вариантами (A_1 и A_2) и, кроме того, сравнить A_1 и A_2 (без учета A_0). Эти два способа разбивки мы обозначим символами A_I и A_II. Остальные два фактора B и C, имеющие лишь по два варианта каждый, можно разбить лишь единственным способом: сравнив B_1 с B_2 и C_1 с C_2. Эти два приема мы обозначим соответственно B и C (без значка). Итак, 4 элементарные разбивки могут быть обозначены здесь следующими заголовками:

(A_I)	(A_II)	(B)	(C)
$+2A_0$	$+A_1$	$+B_1$	$+C_1$
$-\ A_1$	$-A_2$	$-B_2$	$-C_2$
$-\ A_2$			

$J_5 = 25,02$
$J_1 = 45,36$
$J_{01} = 79,74$

			S	A_I	A_{II}	B	C	A_I B	A_{II} B	A_I C	A_{II} C	B C	A_I B C	A_{II} B C	Контроль Σe	Контроль S^2
				$+2A_0$ $-A_1$ $-A_2$	$+A_1$ $-A_2$	$+B_1$ $-B_2$	$+C_1$ $-C_2$									
A_0	B_1	C_1	15,6	+2	0	+1	+1	+2	0	+2	0	+1	+2	0	+11	243,36
		C_2	17,4	+2	0	+1	−1	+2	0	−2	0	−1	−2	0	− 1	302,76
	B_2	C_1	19,2	+2	0	−1	+1	−2	0	+2	0	−1	−2	0	− 1	368,64
		C_2	24,6	+2	0	−1	−1	−2	0	−2	0	+1	+2	0	− 1	605,16
A_1	B_1	C_1	15,6	−1	+1	+1	+1	−1	+1	−1	+1	+1	−1	+1	+ 3	243,36
		C_2	6,6	−1	+1	+1	−1	−1	+1	+1	−1	−1	+1	−1	− 1	43,56
	B_2	C_1	15,6	−1	+1	−1	+1	+1	−1	−1	+1	−1	+1	−1	− 1	243,36
		C_2	3,0	−1	+1	−1	−1	+1	−1	+1	−1	+1	−1	+1	− 1	9,00
A_2	B_1	C_1	3,0	−1	−1	+1	+1	−1	−1	−1	−1	+1	−1	−1	− 5	9,00
		C_2	8,4	−1	−1	+1	−1	−1	−1	+1	+1	−1	+1	+1	− 1	70,56
	B_2	C_1	10,2	−1	−1	−1	+1	+1	+1	−1	−1	−1	+1	+1	− 1	104,04
		C_2	4,8	−1	−1	−1	−1	+1	+1	+1	+1	+1	−1	−1	− 1	23,04
Σe^2				24	8	12	12	24	8	24	8	12	24	8		2 265,84
ΣeS				+86,4	+14,4	−10,8	+14,4	−21,6	+ 7,2	−36,0	+21,6	−10,8	+21,6	+ 7,2	+93,6	
$J = \frac{(\Sigma eS)^2}{\Sigma e^2}$				311,04	25,92	9,72	17,28	19,44	6,48	54,00	58,32	9,72	19,44	6,48	537,84	

Из этих элементарных разбивок можно составить 5 двойных сочетаний:

$$\left|\begin{array}{c|c|c|c|c} (A_{\mathrm{I}}) & (A_{\mathrm{II}}) & (A_{\mathrm{I}}) & (A_{\mathrm{II}}) & (B) \\ (B) & (B) & (C) & (C) & (C) \end{array}\right|$$

и две комбинации по три элемента:

$$\left|\begin{array}{c|c} (A_{\mathrm{I}}) & (A_{\mathrm{II}}) \\ (B) & (B) \\ (C) & (C) \end{array}\right|$$

Всего, стало быть, здесь можно проанализировать $4+5+2=11$ разбивок и сочетаний, т. е. на 1 меньше, чем общее число имеющихся в нашем распоряжении сумм S (это является общим правилом). Само собою разумеется, что в действительных условиях могут понадобиться не все эти 11 комбинаций (здесь они приводятся лишь для общности приема). Вот отвечающая им таблица (см. стр. 62).

Здесь все клетки заполнены по правилам, описанным в применении к предыдущему примеру, но дополнительно включены два контрольных столбца, озаглавленные Σe и S^2.

В первом из них (Σe) числа составлены как сумма всех одиннадцати чисел e, стоящих в тех же строчках в основной части таблицы (с учетом знаков $+$ и $-$). Так, в частности, первое из этих чисел Σe (самое верхнее) получено так:

$$+2+0+1+1+2+0+2+0+1+2+0=11.$$

Аналогично найдены и все остальные числа этого столбца. Общая их сумма очевидно $=0$ (контроль).

Далее с этими числами поступлено так же, как и со всеми числами e в основной части таблицы, т. е. каждое из них помножено на соответствующую сумму S (стоящую слева в той же строчке), и все эти произведения сложены. У нас эта сумма получилась равной $+93,6$ (записана внизу).

Полученная этим способом сумма ($+93,6$) должна в то же время служить и суммой всех 11-ти записанных слева от нее сумм (полученных при умножении отдельных чисел e на те же суммы S с последующим их сложением). В самом деле, сложив все эти суммы (ΣeS), получим:

$+86,4+14,4-10,8+14,4-21,6+7,2-36,0+21,6-10,8+$
$+21,6+7,2=93,6.$

Далее проверяем последние числа J, для чего складываем их:

$\Sigma J=311,04+25,92,+9,72+17,28+19,44+6,48+54,00+58,32+$
$+9,72+19,44+6,48=537,84.$

Эту же сумму ($\Sigma J=587,84$) мы для контроля можем получить и другим путем. Для этого следует возвысить все 12 записанных слева сумм S в квадрат, результаты занести в крайний справа столбец (озаглавленный S^2) и все эти числа сложить. Получим:

$$\Sigma S^2=2265,84.$$

Из этой суммы следует отнять величину, равную, $\dfrac{n(\Sigma S)^2}{N}$.

В нашем примере было $N=36$ дат, которые в каждом сочетании трех факторов повторялись по $n = 3$ раза. Сумма всех наших чисел S (или, что все равно, сумма всех 36-ти первоначальных дат V) у нас $= 144$:

$$\Sigma S = 144.$$

Следовательно, эта поправка:

$$\frac{n(\Sigma S)^2}{N} = \frac{3 \cdot (144)^2}{36} = \frac{20736}{12} = 1728.$$

Вычитая ее из суммы $\Sigma S^2 = 2265{,}84$, получим:

$$\Sigma S^2 - \frac{n(\Sigma S)^2}{N} = 2265{,}84 - 1728 = 537{,}84.$$

Этот результат полностью совпал в ранее вычисленной суммой всех одиннадцати чисел J. Таким образом эта контрольная формула имеет следующий вид:

$$\Sigma J = \Sigma S^2 - \frac{n(\Sigma S)^2}{N}.$$

Данная контрольная формула будет справедлива лишь в том случае, если нами использованы все мыслимые способы разбивки и сочетаний, общее число которых равно числу всех сумм S без единицы.

После проверки всех чисел J приступаем к их оценке, для чего вычисляем три ориентировочные значения:

$$\left.\begin{aligned} J_5 &= n\sigma_z^2\Theta_5 \\ J_1 &= n\sigma_z^2\Theta_1 \\ J_{01} &= n\sigma_z^2\Theta_{01} \end{aligned}\right\} \quad \begin{gathered}\text{при числе степеней} \\ \text{свободы } \nu_1 = 1 \\ \text{и } \nu_2 = \nu_z\end{gathered}$$

у нас $n = 3$ $\sigma_z^2 = 2$ $\nu_z = 30$. Следовательно:

$$n\sigma_z^2 = 3 \cdot 2 = 6.$$

По таблице Фишера-Стьюдента при $\nu_1 = 1$ и $\nu_z = 30$ находим:

$$\Theta_5 = 4{,}17$$
$$\Theta_1 = 7{,}56$$
$$\Theta_{01} = 13{,}29.$$

Следовательно:

$$J_5 = 6 \cdot 4{,}17 = 25{,}02$$
$$J_1 = 6 \cdot 7{,}56 = 45{,}36$$
$$J_{01} = 6 \cdot 13{,}29 = 79{,}76.$$

Эти три ориентировочные числа записаны в левом верхнем углу нашей таблицы. Сравнивая полученные ранее 11 чисел J с этими тремя их значениями, мы видим, что из всех 11-ти испытанных здесь комбинаций достоверными оказались лишь 4.

1) A_0 оказалось существенно отличным от A_1 и A_2 (с вероятностью, превышающей 0,999).

2) Выяснилось, что A_1 также существенно отличается от A_2 (с вероятностью, превышающей 0,95).

3 и 4) Обнаружено существенно эффективное совместное действие факторов A и C, как при отделении A_0 от A_1 и A_2, так и при разбивке этого фактора на A_1 и A_2 (без учета A_0) в обоих случаях с вероятностью, превышающей 0,99.

Остальные различия и соотношения здесь не подтвердились (статистически не доказаны), так как соответствующие им числа J оказались меньше самого малого из трех ориентировочных значений J ($J_5 = 25{,}02$).

XIX

Выше был рассмотрен общий порядок статистической разработки комплекса признаков, отдельные даты которого (V) были собраны по „схеме эксперимента". Для ознакомления с методами статистического анализа комбинационных таблиц, составленных по „схеме наблюдения", приведем предварительно следующий сокращенный числовой пример.

Пусть исследователя интересует вопрос о влиянии отдельных градаций (вариантов) какого-либо не поддающегося заблаговременной дозировке количественного или качественного фактора A на некоторый результативный признак V (например, влияние температуры или влажности на урожай; кровяного давления на функцию какого-либо органа). Предположим, что исследователь сделал $n = 16$ отдельных наблюдений и на основании рассмотрения их результатов разбил все полученные им 16 дат V на группы по величине (или по качественным особенностям) фактора A, причем оказалось, что при первом варианте этого фактора (A) даты V получились равными:

$$5 \text{ и } 1;$$

при следующем его значении A_2:

$$4, 6, 7, 5 \text{ и } 8;$$

в применении к значению A_3:

$$8, 9, 7, 6, 10 \text{ и } 8,$$

а для последнего значения A_4:

$$4, 2 \text{ и } 6.$$

Помещая все эти даты V в клетки соответствующей таблицы (с определенными заголовками по признаку A), получим:

A_1	A_2	A_3	A_4
5 1	4 7 5 6 8	8 7 10 9 6 8	4 6 2

Следовательно, отдельные частные средние:

$$A_1=\frac{5+1}{2}=\frac{6}{2}=3$$

$$A_2=\frac{4+6+7+5+8}{5}=\frac{30}{5}=6$$

$$A_3=\frac{8+9+7+6+10+8}{6}=\frac{48}{6}=8$$

$$A_4=\frac{4+2+6}{3}=\frac{12}{3}=4,$$

а общее среднее:

$$M=\frac{\Sigma V}{n}=\frac{96}{16}=6$$

(так как сумма всех дат $\Sigma V=96$).

Нетрудно убедиться в том, что отклонение каждой даты x от общего среднего M равно сумме двух отклонений: z — отклонения той же даты V от соответствующего частного среднего A и α — отклонения этого частного среднего A от общего среднего M. Например, первая из выписанных выше дат ($V=5$ в группе A_1) отклоняется от $M=6$ на величину:

$$x=V-M=5-6=-1;$$

та же дата ($V=5$) отклоняется от своего частного среднего $A_1=3$ на величину:

$$z=V_1-A_1=5-3=+2,$$

а это частное среднее $A_1=3$ в свою очередь отклоняется от общего среднего $M=6$ на величину:

$$\alpha=A_1-M=3-6=-3.$$

A	V	x	x^2	z	z^2	α	α^2
$A_1=3$	5 1	−1 −5	1 25	+2 −2	4 4	−3 −3	9 9
$A_2=6$	4 6 7 5 8	−2 0 +1 −1 +2	4 0 1 1 4	−2 0 +1 −1 +2	4 0 1 1 4	0 0 0 0 0	0 0 0 0 0
$A_3=8$	8 9 7 6 10 8	+2 +3 +1 0 +4 +2	4 9 1 0 16 4	0 +1 −1 −2 +2 0	0 1 1 4 4 0	+2 +2 +2 +2 +2 +2	4 4 4 4 4 4
$A_4=4$	4 2 6	−2 −4 0	4 16 0	0 −2 +2	0 4 4	−2 −2 −2	4 4 4
Сумма	96	0	90	0	36	0	54

Отсюда видно, что сумма:

$$z + \alpha = +2 - 3 = -1$$

в точности совпала с величиной $x = -1$.
Итак имеем равенство:

$$x = z + \alpha.$$

Вычислим все эти три отклонения (x, z и α) для всех $n = 16$ наших дат V, возвысим их в квадрат и сложим (см. таблицу на стр. 66 внизу).

Суммы квадратов этих отклонений мы обозначим буквами S с соответствующими значками. У нас, следовательно:

$$S_x = 90 \qquad S_z = 36 \qquad S_a = 54.$$

Нетрудно убедиться в том, что:

$$S_x = S_z + S_a.$$

Число степеней свободы общего рассеяния:

$$\nu_x = n - 1 = 16 - 1 = 15.$$

Число вариантов фактора A у нас $= 4$ ($a = 4$). Следовательно, для суммы S_a:

$$\nu_a = a - 1 = 4 - 1 = 3.$$

Откуда число степеней свободы для суммы S_z (как остаток):

$$\nu_z = \nu_x - \nu_a = 15 - 3 = 12.$$

Разделив S на ν, получим квадраты мер варьирования (σ). Для характеристики колеблемости наших дат V под действием фактора A получим:

$$\sigma_a^2 = \frac{S_a}{\nu_a} = \frac{54}{3} = 18,$$

а для характеристики их колеблемости под влиянием всех остальных (не учтенных в опыте „случайных") воздействий:

$$\sigma_z^2 = \frac{S_z}{\nu_z} = \frac{36}{12} = 3.$$

Отсюда показатель:

$$\Theta = \frac{\sigma_a^2}{\sigma_z^2} = \frac{18}{3} = \underline{\underline{6}} \text{ (при } \nu_1 = 3 \text{ и } \nu_2 = 12).$$

Следовательно, влияние фактора A на результативный признак V можно считать статистически доказанным с вероятностью, превышающей $P = 0{,}99$.

Для сравнения отдельных частных средних друг с другом следовало бы вычислить квадрат средних ошибок каждого из них по формулам типа:

$$m^2 = \frac{\sigma_z^2}{p},$$

где p — численность той группы дат V, для которых было получено данное частное среднее.

Средняя ошибка разности двух таких средних (напр. $A_3 - A_2$):

$$\Delta = \sqrt{m_3^2 + m_2^2} = \sqrt{\frac{\sigma_z^2}{p_3} + \frac{\sigma_z^2}{p_2}}$$

В нашем примере:

для $A_3 = 8$ $\quad p_3 = 6 \quad \sigma_z^2 = 3,$
для $A_2 = 6$ $\quad p_2 = 5$

следовательно:

$$\Delta = \sqrt{\frac{3}{6} + \frac{3}{5}} = \sqrt{0,5 + 0,6} = \sqrt{1,10} = 1,05.$$

Для оценки достоверности различия между A_3 и A_2 вычисляем отношение:

$$t = \frac{A_3 - A_2}{\Delta} = \frac{8-6}{1,05} = \frac{2}{1,05} = 1,91.$$

При $\nu = \nu_z = 12$ находим, что такое значение t не свидетельствует о достоверности этой разности (статистически не доказано).

Аналогично можно было бы поступить и с остальными разностями.

XX

Покажем теперь другой прием получения σ_a^2, σ_z^2, Θ и t, более удобный в применении к анализу совокупного действия нескольких признаков. Составим таблицу из такого числа клеток, сколько группировок было отмечено по отдельным факторам A (у нас, следовательно, из 4 клеток), поставим около этих клеток соответствующие заголовки (A_1, A_2, A_3 и т. д.) и внутри каждой такой клетки напишем 4 числа: в левом верхнем углу поставим число (частоту) p, показывающее, сколько дат V оказалось в данной их группе; под ним в левом нижнем углу той же клетки запишем общую сумму этих дат ΣV; в правом верхнем углу клетки поставим соответствующее частное среднее (сумма дат, деленная на их частоту p), а в нижнем правом углу запишем новое число (обозначим его буквой k), получаемое или путем деления квадрата суммы дат (ΣV) на число этих дат p:

$$k = \frac{(\Sigma V)^2}{p}$$

или, что то же, путем умножения квадрата вышестоящего среднего (A) на частоту p:

$$k = pA^2,$$

или же, наконец, как произведение среднего A на сумму ΣV:

$$k = A \Sigma V$$

A_1		A_2		A_3		A_4	
2	3	5	6	6	8	3	4
6	18	30	180	48	384	12	48

Так, в частности, для первой по порядку группы дат V, соответствующих первому варианту фактора A_1, т. е. для двух дат V:

$$5 \text{ и } 1$$

мы имеем $p = 2$ $\quad \Sigma V = 5 + 1 = 6$, частное среднее:

$$A = \frac{\Sigma V}{p} = \frac{6}{2} = 3,$$

а число:

$$k = \frac{(\Sigma V)^2}{p} = \frac{6^2}{2} = \frac{36}{2} = 18$$

(или $k = pA_1^2 = 2 \cdot 3^2 = 2 \cdot 9 = 18$ или $k = A\Sigma V = 3.6 = 18$).

Эти четыре числа (2, 6, 3 и 18) и записаны в первой клетке приведенной выше таблицы. Аналогично вычислены эти данные и для всех остальных клеток.

Чтобы не связывать себя определенным фактором, напр. A (так как в дальнейшем мы будем учитывать также и комбинации нескольких факторов), обозначим общее число сделанных группировок буквой g (а не буквой a). У нас следовательно:

$$g = 4.$$

Сумма всех чисел p равна очевидно общему числу всех дат V, т. е.:

$$n = \Sigma p = 2 + 5 + 6 + 3 = 16.$$

Сумма всех частных сумм дат, записанных в отдельных клетках, равна общей сумме всех $n = 16$ дат V, т. е.:

$$\Sigma V = 6 + 30 + 48 + 12 = 96.$$

С помощью этой общей суммы (96) и общего количества всех дат (16) определим число k (назовем его k_0) также и для этих итогов:

$$k_0 = \frac{(\Sigma V)^2}{n} = \frac{(96)^2}{16} = \frac{9216}{16} = 576.$$

Заметим, что это число k_0 не совпадает с суммой всех чисел k, полученных ранее в отдельных клетках нашей таблицы:

$$\Sigma k = 18 + 180 + 384 + 48 = 630.$$

Для применения новых формул нам необходимо еще подсчитать сумму квадратов всех 16 дат V. В нашем примере:

$$\Sigma V^2 = 666.$$

С помощью этих данных мы можем теперь вычислить все ранее полученные показатели по следующим новым формулам:

$$\sigma_a^2 = \frac{\Sigma k - k_0}{g - 1} = \frac{630 - 576}{4 - 1} = \frac{54}{3} = 18$$

$$\sigma_z^2 = \frac{\Sigma V^2 - \Sigma k}{n - g} = \frac{666 - 630}{16 - 4} = \frac{36}{12} = 3,$$

откуда показатель:

$$\Theta = \frac{\sigma_a^2}{\sigma_z^2} = \frac{18}{3} = 6 \text{ при числе степеней свободы:}$$

$$\nu_1 = g - 1 = 4 - 1 = 3 \text{ и}$$

$$\nu_2 = n - g = 16 - 4 = 12.$$

Для оценки достоверности различий между отдельными парами частных средних (A_1, A_2, A_3 и т. д.) удобнее вычислять не отношение t, а его квадрат t^2, совпадающий в таблице Фишера-Стьюдента со значением Θ для $\nu_1 = 1$ (т. е. с числами, помещенными там в крайнем слева столбце). При сравнении двух частных средних (обозначим их в общем виде M_1 и M_2, чтобы не связывать себя ни с каким определенным фактором) это значение вычисляется по формуле:

$$\Theta = \left(\frac{p_1\, p_2}{p_1 + p_2}\right) \cdot \frac{(M_1 - M_2)^2}{\sigma_z^2}$$

при числе степеней свободы:

$$\nu_1 = 1 \qquad \nu_2 = n - g.$$

Пусть попрежнему требуется оценить достоверность различия между третьим и вторым частным средним, т. е. между средними 8 и 6, вычисленными соответственно из 6 и из 5 дат. Здесь следовательно:

$$M_1 = 8 \qquad p_1 = 6$$
$$M_2 = 6 \qquad p_2 = 5,$$

откуда (при $\sigma_z^2 = 3$) искомое значение:

$$\Theta = \left(\frac{6 \cdot 5}{6+5}\right)\frac{(8-6)^2}{3} = \frac{30}{11} \cdot \frac{4}{3} = \frac{40}{11} = 3{,}64.$$

В крайнем слева столбце таблицы Фишера-Стьюдента для значения $\nu_1 = 1$ ищем клетку, соответствующую числу степеней свободы:

$$\nu_2 = n - g = 16 - 4 = 12.$$

В этой клетке стоят три числа:

4,75
9,33
18,64.

Наше значение $\Theta = 3{,}64$ оказалось меньше всех этих чисел, следовательно, разность этих средних здесь не доказана.

Ранее при оценке этой же разности мы получили значение $t = 1{,}91$, квадрат которого:

$$t^2 = (1{,}91)^2 = 365$$

приблизительно совпал с этим значением:

$$\Theta = 3{,}64$$

(разница объясняется потерей точности при округлении в процессе вычислений).

XXI

После этих предварительных замечаний рассмотрим следующий числовой пример, иллюстрирующий влияние трех факторов. A, B и C на количественный результативный признак V:

		C_1	C_2	C_3	C_4
A_1	B_1				
	B_2				9
	B_3		5		7 5 3
A_2	B_1				
	B_2	1	4 6 6 4	7 5 5 7	
	B_3				1
A_3	B_1	2 0 0 2	2 4 4 3 2	1	
	B_2				
	B_3				

Здесь датами V заполнены не все клетки, и в отдельных местах скопилось не одинаковое количество дат. Всего здесь имеется:

$$n = 25 \text{ дат},$$

общая сумма которых:

$$\Sigma V = 95,$$

а сумма квадратов:

$$\Sigma V^2 = 501$$

(в чем легко убедиться непосредственным вычислением).

Эта таблица является черновой, и ее следует преобразовать так, чтобы в каждой клетке были записаны указанные выше 4 числа: количество дат, их сумма, средняя величина этих дат и число k (равное квадрату суммы дат, деленному на их количество; или квадрату среднего, умноженного на частоту; или среднему, умноженному на сумму дат). Вот эта преобразованная таблица:

		C_1	C_2	C_3	C_4
A_1	B_1				
	B_2				1 9 9 81
	B_3		1 5 5 25		3 5 15 75
A_2	B_1				
	B_2	1 1 1 1	4 5 20 100	4 6 24 144	
	B_3				1 1 1 1
A_3	B_1	4 1 4 4	5 3 15 45	1 1 1 1	
	B_2				
	B_3				

Здесь мы имеем 10 заполненных клеток (остальные 26 пустые). Следовательно, число фактически получившихся группировок наших $n = 25$ дат по комбинациям вариантов всех трех факторов (A, B и C):

$$g = 10.$$

Сложив в этих клетках все числа k (стоящие в правых нижних углах), получим:

$$\Sigma k = 81 + 25 + 75 + 1 + 100 + 144 + 1 + 4 + 45 + 1 = 477.$$

А для всей таблицы в целом (по итогам):

$$k_0 = \frac{(\Sigma V)^2}{n} = \frac{(95)^2}{25} = \frac{9025}{25} = 361.$$

Следовательно, квадрат меры варьирования, характеризующей колеблемость всех наших дат V под влиянием воздействий со стороны всех трех факторов A, B, C одновременно:

$$\sigma^2_{abc} = \frac{\Sigma k - k_0}{g - 1} = \frac{477 - 361}{10 - 1} = \frac{116}{9} = 12{,}89,$$

а квадрат меры варьирования, характеризующей колеблемость тех же дат V под влиянием всех остальных, неучтенных в опыте („случайных") факторов:

$$\sigma_z^2 = \frac{\Sigma V^2 - \Sigma k}{n - g} = \frac{501 - 477}{25 - 10} = \frac{24}{15} = 1{,}60.$$

Далее вычисляем показатель:

$$\Theta_{abc} = \frac{\sigma^2_{abc}}{\sigma_z^2} = \frac{12{,}89}{1{,}60} = \underline{\underline{8{,}06}}$$

при: $\quad \nu_1 = g - 1 = 10 - 1 = 9$

и $\quad \nu_2 = n - g = 25 - 10 = 15.$

Следовательно, совместное действие на результативный признак V со стороны трех изучаемых факторов A, B и C (по сравнению с влиянием на них со стороны всех остальных не учтенных в опыте факторов) здесь можно считать статистически доказанным с вероятностью, превышающей 0,999.

Для сравнения между собой какой-либо пары частных средних (записанных в правом углу соответствующих клеток), например: среднего $M_1 = 5$, полученного из $p_1 = 3$ дат, возникших под действием комбинации факторов A_1, B_3, C_4 и среднего $M_2 = 3$, полученного из $p_2 = 5$ дат, соответствующих сочетанию A_3, B_1, C_2, получим (при $\sigma_z^2 = 1{,}60$):

$$\Theta = \left(\frac{p_1\,p_2}{p_1 + p_2}\right)\frac{(M_1 - M_2)^2}{\sigma_z^2} = \left(\frac{3 \cdot 5}{3 + 5}\right)\frac{(5 - 3)^2}{1{,}6} =$$

$$= \frac{15}{8} \cdot \frac{4}{1{,}6} = \underline{4{,}69}$$

при $\nu_1 = 1$ и $\nu_2 = n - g = 25 - 10 = 15.$

Следовательно, сочетание A_1, B_3, C_4, давшее результат $M_1=5$, является в каком-то отношении более выгодным по сравнению с комбинацией A_3, B_1, C_2, в результате которой возникло среднее $M_2=3$, причем разница между этими средними может считаться статистически доказанной с вероятностью, превышающей 0,95.

Аналогично можно было бы выяснить достоверность различия между остальными парами этих частных средних (по мере действительной в том надобности).

Из полной таблицы сочетаний всех трех факторов легко получить ряд частных таблиц, характеризующих результаты совместного действия одних факторов без учета других, например, только факторов A и C без учета фактора B. Вот как будет выглядеть такая таблица:

	C_1	C_2	C_3	C_4
A_1		1 5 5 25		4 6 24 144
A_2	1 1 1 1	4 5 20 100	4 6 24 144	1 1 1 1
A_3	4 1 4 4	5 3 15 45	1 1 1 1	

Здесь частота p и сумма дат ΣV (т. е. числа, стоящие в каждой клетке слева) получены путем простой суммации соответствующих чисел полной таблицы при объединении нескольких клеток в одну, а частные средние и числа k (стоящие справа) заново вычислены в каждой клетке этой новой таблицы. Так, например, в полной таблице сочетанию A_1 и C_4 соответствовали три клетки (для B_1, B_2 и B_3) с такими записями:

для B_1	
для B_2	1 9 9 81
для B_3	3 5 15 75

Объединяя эти клетки в одну, мы должны будем сложить две частоты p:

$$1+3=4$$

и обе суммы ΣV:

$$9+15=24$$

(но не суммировать справа стоящие числа: $9+5$ и $81+75$).

В результате получим новую объединенную клетку с такими записями:

4	6
24	144

в которой справа стоят заново (для этой клетки) вычисленные показатели: среднее

$$\frac{24}{4}=6$$

и число:

$$k=\frac{(24)^2}{4}=\frac{576}{4}=144.$$

Аналогично заполнены и остальные клетки этой новой таблицы.

Здесь общее число группировок:

$$g=9,$$

а сумма всех чисел k:

$$\Sigma k=25+144+1+100+144+1+4+45+1=465,$$

откуда:

$$\sigma_{ac}^2=\frac{\Sigma_k-k_a}{g-1}=\frac{465-361}{9-1}=\frac{104}{8}=13$$

$$\sigma_z^2=\frac{\Sigma V^2-\Sigma k}{n-g}=\frac{501-465}{25-9}=\frac{36}{16}=2{,}25$$

(так как n, ΣV^2 и k_0 всюду остаются неизменными).

Следовательно, показатель:

$$\Theta_{ac}=\frac{\sigma_{ac}^2}{\sigma_z^2}=\frac{13{,}00}{2{,}25}=\underline{\underline{5{,}76}}$$

при:

$$\nu_1=g-1=9-1=8$$
$$\nu_2=n-g=25-9=16.$$

Желая исследовать влияние на тот же результативный признак V со стороны комбинации вариантов фактора B и C одновременно, получим таблицу:

	C_1	C_2	C_3	C_4
B_1	4 1 4 4	5 3 15 45	1 1 1 1	
B_2	1 1 1 1	4 5 20 100	4 6 24 144	1 9 9 81
B_3		1 5 5 25		4 4 16 64

Здесь $\Sigma k = 465$ и $g = 9$ (попрежнему), следовательно, получим, как и раньше:

$$\sigma_{bc}^2 = 13 \qquad \sigma_z^2 = 2{,}25 \qquad \Theta_{bc} = \underline{\underline{5{,}76}}$$

Для сочетания факторов A и B будем иметь таблицу:

	B_1	B_2	B_3
A		1 9 9 81	4 5 20 100
A_2		9 5 45 225	1 1 1 1
A_3	10 2 20 40		

$$\Sigma k = 447 \qquad g = 5$$

откуда:

$$\sigma_{ab}^2 = \frac{447 - 361}{5 - 1} = \frac{86}{4} = 21{,}5$$

$$\sigma_z^2 = \frac{501 - 447}{25 - 5} = \frac{54}{20} = 2{,}7$$

при:

$$\Theta_{ab} = \frac{21{,}5}{2{,}7} = \underline{\underline{7{,}96}}$$

$$\nu_1 = 5 - 1 = 4$$

$$\nu_2 = 25 - 5 = 20$$

Из этих трех малых таблиц можно сделать аналогичную выборку для анализа действия на результативный признак со стороны только какого-либо одного фактора, напр. A:

A_1	5 29	5,8 168,2
A_2	10 46	4,6 211,6
A_3	10 20	2 40

Здесь:

$$\Sigma k = 419{,}8 \qquad g = 3$$

и следовательно:

$$\sigma_a^2 = 29{,}4 \qquad \sigma_z^2 = 3{,}69$$

$$\Theta_a = \underline{\underline{7{,}97}}$$

при:

$$\nu_1 = 2 \qquad \nu_2 = 22$$

В применении к фактору B:

B_1	10 20	2 40
B_2	10 54	5,4 291,6
B_3	5 21	4,2 88,2

$$\Sigma k = 419{,}8 \qquad g = 3$$

(данные совпали с предыдущими):

$$\Theta_b = \underline{\underline{7{,}97}}$$

Для фактора C будем иметь:

C_1	5 5	1 5
C_2	10 40	4 160
C_3	5 25	5 125
C_4	5 25	5 125

$$\Sigma k = 415 \qquad g = 4$$

$$\sigma_c^2 = 18{,}0 \qquad \sigma_z^2 = 4{,}1$$

$$\Theta_c = 4{,}39$$

при:

$$\nu_1 = 3 \qquad \nu_2 = 21.$$

Если бы мы захотели сравнить любую пару средних, напр. $M_1 = 5$ (для группировки C_3) при $p_1 = 5$ и $M_2 = 4$ (для группировки C_2) при $p_2 = 10$, то получили бы значение:

$$\Theta = \frac{p_1 \cdot p_2}{p_1 + p_2} \frac{(M_1 - M_2)^2}{\sigma_z^2} = \left(\frac{5 \,.\, 10}{5 + 10}\right) \frac{(5-4)^2}{4{,}1} = \frac{50}{15} \cdot \frac{1}{4{,}1} = 0{,}81$$

при $\nu_1 = 1$ $\qquad \nu_2 = n - g = 25 - 4 = 21$

(явно не достоверное, так как $\Theta < 1$).

Итак выяснилось, что все три изучаемые фактора A, B и C как порознь, так и совместно оказывают в нашем примере несомненное влияние на результативный признак V, так как все 7 показателей:

$$\Theta_a = 7{,}97 \qquad \Theta_{ab} = 7{,}96$$

$$\Theta_b = 7{,}97 \qquad \Theta_{ac} = 5{,}76 \qquad \Theta_{abc} = 8{,}06$$

$$\Theta_c = 4{,}39 \qquad \Theta_{bc} = 5{,}76$$

удовлетворяют здесь условию достоверности (Θ_c гарантирует вероятность 0,95, Θ_{abc} — вероятность 0,999, а все остальные — вероятность 0,99).

XXII

Вычисленные показатели Θ имеют лишь служебное значение, позволяя использовать критерии оценки достоверности по таблице Фишера-Стьюдента, сами же по себе не обладают особой наглядностью (один и тот же показатель Θ при различном числе степеней свободы ν_1 и ν_2 будет иметь неодинаковое значение). Покажем способ определения „мер связи“, характеризующих степень зависимости („тесноту связи“) результативного признака V от одного или нескольких изучаемых факторов (A, B, C и т. д.).

Вернемся для этого к разобранному уже выше краткому (вводному) числовому примеру, иллюстрирующему действие $g = 4$ вариантов (дозировок) какого-либо признака A на величину $n = 16$ дат V:

A_1	A_2	A_3	A_4
5 1	4 7 5 6 8	8 7 10 9 6 8	4 2 6

Вспомним, что в применении к данному примеру сумма квадратов отклонений (x) всех этих дат V от их общего среднего $M=6$:

$$S_x = 90.$$

Сумма квадратов отклонений (z) тех же дат V от их частных средних ($A_1 = 3$, $A_2 = 6$, $A_3 = 8$ и $A_4 = 4$):

$$S_z = 36$$

Сумма же квадратов отклонений (α) этих частных средних (A_1, A_2 и т. д.) от общего среднего $M=6$ (подсчитанная для всех $n=16$ дат):

$$S_a = 54.$$

Вспомним далее, что:

$$S_x = S_z + S_a$$

$$90 = 36 + 54.$$

Здесь сумма $S_x = 90$ характеризует общее рассеяние всех дат V около $M=6$, фактически обнаруженное в наших наблюдениях; сумма $S_a = 54$ измеряет ту часть колеблемости этих дат V, которая вызывается действием фактора A, а сумма $S_z = 36$ зависит только от рассеяния наших дат под действием всех остальных не учтенных нами („случайных“) влияний.

Эффективность фактора A (в его действии на признак V) характеризуется суммой S_a, которая у нас оказалась $= 54$. Спрашивается, какого максимального предела могла бы достичь эта сумма S_a (при максимальной эффективности фактора A)? Очевидно это имело бы место тогда, когда никаких других факторов (кроме A) не существовало бы вовсе, т. е. при:

$$S_z = 0$$

и тогда, следовательно, сумма:

$$S_x = S_z + S_a = 0 + S_a = S_a.$$

Итак, максимальное значение суммы S_a:

$$\max S_a = S_x.$$

Получив какое-либо частное значение этой суммы (напр. $S_a = 54$), мы можем сравнить его с этим предельным значением (т. е. с суммой $S_x = 90$) и ответить на вопрос: какую часть данная сумма (54) составляет от своего возможного максимума (90):

$$\frac{S_a}{S_x} = \frac{54}{90} = 0{,}6.$$

Это отношение (0,6) и принимается за искомую меру связи (ее квадрат), обозначаемую обычно греческой буквой η (эта) и называемую „корреляционным отношением“. Итак, у нас:

$$\eta^2 = 0{,}6 \qquad \eta = \sqrt{0{,}60} = 0{,}77.$$

Корреляционное отношение можно получить и с помощью ранее вычисленных мер варьирования σ. Вспомним, что в нашем примере:

$$\text{(при } \nu_x = 15 \qquad \nu_2 = 3 \qquad \text{и} \qquad \nu_z = 12\text{):}$$

$$\sigma_x^2 = \frac{S_x}{\nu_x} = \frac{90}{15} = 6 \qquad \sigma_a^2 = \frac{S_a}{\nu_a} = \frac{54}{3} = 18$$

$$\sigma_z^2 = \frac{S_z}{\nu_z} = \frac{36}{12} = 3.$$

По этим данным корреляционное отношение (его квадрат) можно получить по следующей окончательной формуле:

$$\eta^2 = \frac{(g-1)\,\sigma_a^2}{(g-1)\,\sigma_a^2 + (n-g)\,\sigma_z^2}$$

В нашем примере (при $g = 4$ и $n = 16$):

$$\eta^2 = \frac{(4-1)\,.\,18}{(4-1)\ 18 + (16-4)\ 3} = \frac{3\,.\,18}{3\,.\,18 + 12\,.\,3} = \frac{54}{54+36} = \frac{54}{90} = 0{,}6,$$

что мы имели и раньше.

Если бы данный фактор (A) вовсе не влиял на признак V (не вызывал бы никаких его колебаний), то:

$$\sigma_a^2 = 0$$

и стало быть:

$$\eta^2 = 0;$$

если же, наоборот, все остальные факторы (кроме A) не порождали бы никакого варьирования дат V, то:

$$\sigma_z^2 = 0,$$

а показатель:

$$\eta^2 = 1.$$

Следовательно, корреляционное отношение η всегда заключается в пределах от 0 до 1.

Однако практически показатель η никогда не бывает $= 0$, так как все частные средние, получаемые по отдельным группировкам фактора A, всегда обнаруживают некоторые минимальные (случайные) колебания, мера которых (σ_a^2) в силу этого в действительности никогда не обращается в 0. Если данный фактор A никакого преимущественного действия на признак V не имеет и колеблет даты V наравне со всеми остальными неучтенными в опыте („случайными“) факторами, то мера колеблемости σ_a^2 совпадает с мерой случайных колебаний σ_z^2 и такое ее значение:

$$\sigma_a^2 = \sigma_z^2$$

практически будет минимальным.

При этих условиях получаем некоторое наименьшее значение η (обозначим его η_0), которое соответствует отсутствию закономерной связи между A и V:

$$\eta_0^2 = \frac{g-1}{n-1}.$$

(Эта формула непосредственно получается из предыдущей, если в ней положить $\sigma_a^2 = \sigma_z^2$).

В нашем примере:

$$\eta_0^2 = \frac{4-1}{16-1} = \frac{3}{15} = 0{,}2$$

и следовательно:

$$\eta_0 = \sqrt{0{,}20} = 0{,}45.$$

Итак, при интерпретации показателя η следует учитывать, что минимальное его значение будет не 0, а $\eta_0 = 0{,}45$.

Будем временно оперировать квадратами этих показателей ($\eta^2 = 0{,}6$ и $\eta_0^2 = 0{,}2$).

Наш фактически полученный показатель связи (0,6) превышает минимальный (0,2) на величину разности:

$$\eta^2 - \eta_0^2 = 0{,}6 - 0{,}2 = 0{,}4.$$

Эта величина ($\eta^2 - \eta_0^2$) достигает своего максимума при $\eta^2 = 1$, т. е. тогда, когда она оказывается равной:

$$1 - \eta_0^2 = 1 - 0{,}2 = 0{,}8.$$

Сравнивая данное ее значение (0,4) с этим максимальным (0,8), получим отношение:

$$\frac{\eta^2 - \eta_0^2}{1 - \eta_0^2} = \frac{0{,}4}{0{,}8} = 0{,}5,$$

которое и может служить исправленной мерой связи фактора A с результативным признаком V. Обозначая этот показатель связи буквой ω (омега), получим формулу:

$$\omega^2 = \frac{\eta^2 - \eta_0^2}{1 - \eta_0^2} = 0{,}5,$$

откуда $\omega = \sqrt{0{,}50} = 0{,}71$.

Этот же исправленный показатель связи можно было бы получить и непосредственно по формуле:

$$\omega_1^2 = \frac{(g-1)\ (\sigma_a^2 - \sigma_z^2)}{(g-1)\ \sigma_a^2 + (n-g)\ \sigma_z^2}$$

или по формуле:

$$\omega_2^2 = \frac{\eta^2 (n-1) - g + 1}{n-g}$$

В нашем примере следовательно:

$$\omega^2_1 = \frac{(4-1)\ (18-3)}{(4-1)\ 18 + (16-4)\ .\ 3} = \frac{3\ .\ 15}{3\ .\ 18 + 12\ .\ 3} = \frac{45}{54+36} = \frac{45}{90} = 0{,}5$$

$$\omega^2_2 = \frac{0{,}6\ (16-1) - 4 + 1}{16-4} = \frac{0{,}6\ .\ 15 - 3}{12} = \frac{9-3}{12} = \frac{6}{12} = 0{,}5,$$

что мы имели и раньше.

Полученные выше показатели Θ можно было бы вычислить с помощью этих двух показателей связи η и ω по формулам:

$$\Theta = \frac{\eta^2 (n-g)}{(1-\eta^2)(g-1)} = \frac{0,6\ (16-4)}{(1-0,6)(4-1)} = \frac{0,6 \cdot 12}{(0,4) \cdot 3} = \frac{7,2}{1,2} = 6$$

$$\Theta = \frac{(n-g)\ \omega^2 + (g-1)}{(1-\omega^2)(g-1)} = \frac{(16-4) \cdot 0,5 + (4-1)}{(1-0,5)(4-1)} = \frac{12 \cdot 0,5 + 3}{0,5 \cdot 3} =$$

$$= \frac{6+3}{1,5} = \frac{9}{1,5} = 6.$$

В свою очередь показатели связи η^2 и ω^2 можно получить с помощью найденного ранее отношения Θ по формулам:

$$\eta^2 = \frac{\Theta (g-1)}{\Theta (g-) + (n-g)} \qquad \omega^2 = \frac{(\Theta-1)(g-1)}{\Theta (g-1) + (n-g)}$$

В нашем примере (при $\Theta = 6$, $n = 16$ и $g = 4$):

$$\eta^2 = \frac{6\ (4-1)}{6\ (4-1) + (16-4)} = \frac{18}{18+12} = \frac{18}{30} = = 0,6$$

$$\omega^2 = \frac{(6-1)(4-1)}{6\ (4-1) + (16-4)} = \frac{15}{18+12} = \frac{15}{30} = 0,5,$$

что мы имели и раньше.

В применении ко второму примеру анализа связи трех факторов A, B и C с результативным признаком V (при $n = 25$ наблюдениям) получим:

$$\begin{cases} \eta_a = \sqrt{0,42} = 0,65 \\ \omega_a = \sqrt{0\ 37} = 0,61 \end{cases} \qquad \begin{cases} \eta_{bc} = \sqrt{0,74} = 0,86 \\ \omega_{bc} = \sqrt{0,61} = 0,78 \end{cases}$$

$$\begin{cases} \eta_b = \sqrt{0,42} = 0,65 \\ \omega_b = \sqrt{0,37} = 0,61 \end{cases} \qquad \begin{cases} \eta_{bc} = \sqrt{0,74} = 0,86 \\ \omega_{bc} = \sqrt{0,61} = 0,78 \end{cases}$$

$$\begin{cases} \eta_c = \sqrt{0,39} = 0,62 \\ \omega_c = \sqrt{0,31} = 0,56 \end{cases} \qquad \begin{cases} \eta_{abc} = \sqrt{0,83} = 0,91 \\ \omega_{abc} = \sqrt{0,73} = 0,85 \end{cases}$$

$$\begin{cases} \eta_{ab} = \sqrt{0,61} = 0,78 \\ \omega_{ac} = \sqrt{0,53} = 0,73 \end{cases}$$

Заметим, что исправленные показатели связи ω являются сравнимыми при любом числе группировок g, в то время как на корреляционном отношении η число группировок отражается довольно сильно.

(Это имеет большое значение при сравнении показателей множественной связи с показателями парной связи, так как первые обычно вычисляются для большего числа группировок g, а вторые — при малых значениях g).

Заметим, что при $\sigma_a^2 < \sigma_z^2$ показатель $\omega^2 < 1$, и следовательно ω является мнимой величиной.

+

ν_2 \ ν_1	1	2	3	4	5
17	4,45 8,40 15,72	3,59 6,11 10,66	3,20 5,18 8,73	2,96 4,67 7,68	
18	4,41 8,28 5,38	3,55 6,01 10,39	3,16 5,09 8,49	2,93 4,58 7,46	
19	4,38 ,18 ,08	3,52 5,93 10,16	3,13 5,01 8,28	2,90 4,50 7,26	
20	35 10 82	3,49 5,85 9,95	3,10 4,94 8,10	2,87 4,43 7,10	
21	2 2 2	3,47 5,78 9,77	3,07 4,87 7,94	2,84 4,37 6,95	
22	1	3,44 5,72 9,61	3,05 4,82 7,80	2,82 4,31 6,81	
23	14	3,42 5,66 9,46	3,03 4,76 7,67	2,80 4,26 6,70	
24	4, 7,	3,40 5,61	3,01 4,72	2,78 4,22	

Полученные выше показатели Θ можно было бы вычислить с помощью этих двух показателей связи η и ω по формулам:

	6	8	12	24	∞	t
2,81 4,34 7,02	2,70 4,10 6,56	2,55 3,79 5,96	2,38 3,45 5,32	2,19 3,08 4,63	1,96 2,65 3,85	2,11 2,90 3,96
2,77 4,25 6,81	2,06 4,01 6,35	2,51 3,71 5,76	2,34 3,37 5,13	2,15 3,01 4,45	1,92 2,57 3,67	2,10 2,88 3,92
2,74 4,17 6,61	2,63 3,94 6,18	2,48 3,63 5,59	2,31 3,30 4,97	2,11 2,92 4,29	1,88 2,49 3,52	2.09 2,86 3.88
2,71 4,10 6,46	2,60 3,87 6,02	2,45 3,56 2,44	2,28 3,23 4,82	2,08 2,86 4,15	1,84 2,42 3,38	2,09 2,84 3,85
2,68 4,04 6,32	2,57 3,81 5,88	2,42 3,51 5,31	2,25 3,17 4,70	2,05 2,80 4,03	1,82 2,36 3,26	2,08 2,83 3,82
2,66 3,99 6,19	2,55 3,75 5,76	2,40 3,45 5,19	2,23 3,12 4,58	2,03 2,75 3,92	1,78 2,30 3,15	2,07 2,82 3,79
2,64 3,94 6,08	2,53 3,71 5,56	2,38 3,41 5,09	2,20 3,07 4,48	2,00 2,70 3,82	1,76 2,26 3,05	2,07 2,81 3,77
2,62 3,90	2,51 3,67	2,36 3,36	2,18 3,03	1,98 2,66	1,73 2,21	2,06 2,80

Цена 4 руб.

Таблица R. A. Fisher'a-Student'a

(в переработке по Н. Ф. Деревицкому и G. W. Snedecor'y)

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞	t
1	161,40 4 052,10 406 523,00	199,50 4 999,03 500 016,00	215,72 5 403,49 536 700,00	224,57 5 625,14 562 527,00	230,17 5 764,04 576 449,00	233,97 5 859,39 585 953,00	238,89 5 981,34 598 149,00	243,91 6 105 83 610 598,00	249,04 6 234,16 623 432,00	54,32 656,48 63635,00	12,71 63,66 637,59
2	18,51 98,49 998,46	19,00 99,01 999,00	19,16 99,17 999,20	19,25 99,25 999,20	19,30 99,30 999,20	19,33 99,33 999,20	19,37 99,36 999,40	19,41 99,42 999,60	19,45 99,46 999,40	19,50 99,50 999,40	4,30 9,92 31,60
3	10,13 34,12 167,47	9,55 30,81 148,51	9,28 29,46 141,10	9,12 28,71 137,10	9,01 28,24 134,60	8,94 27,91 132,90	8,84 27,49 130,60	8,74 27,05 128,30	8,64 26,60 125,90	8,53 26,12 123,50	3,18 5,84 12,94
4	7,71 21,20 74,13	6,94 18,00 61,24	6,59 16,69 56,18	6,39 15,98 53,43	6,26 15,52 51,71	6,16 15,21 50,52	6,04 14,80 49,00	5,91 14,37 47,41	5,77 13,93 45,77	5,63 13,46 44,05	2,78 4,60 8,61
5	6,61 16,26 47,04	5,79 13,27 36,61	5,41 12,06 33,20	5,19 11,39 31,09	5,05 10,97 29,75	4,95 10,67 28,83	4,82 10,27 27,64	4,68 9,89 26,42	4,53 9,47 25,14	4,36 9,02 23,78	2,57 4,03 6,86
6	5,99 13,74 35,51	5,14 10,92 26,99	4,76 9,78 23,70	4,53 9,15 21,90	4,39 8,75 20,81	4,28 8,47 20,03	4,15 8,10 19,03	4,00 7,72 17,99	3,84 7,31 16,89	3,67 6.88 15,75	2,42 3,71 5,96
7	5,59 12,25 29,22	4,74 9,55 21,69	4,35 8,45 18,77	4,12 7,85 17,19	3,97 7,46 16,21	3,87 7,19 15,52	3,73 6,84 14,63	3,57 6,47 13,71	3,41 6,07 12,73	3,23 5,65 11,70	2,36 3,50 5,31
8	5,32 11,26 25,42	4,46 8,65 18,49	4,07 7,59 15,83	3,84 7,01 14,39	3,69 6,63 13,49	3,58 6,37 12,86	3,44 6,03 12,04	3,28 5,67 11,19	3,12 5,28 10,30	2 93 4,86 9,35	2,31 3,36 5,04
9	5,12 10,56 22,86	4,26 8,02 16,39	3,86 6,99 13,90	3,63 6,42 12,56	3,48 6,06 11,71	3,37 5,80 11,13	3,23 5,47 10,37	3,07 5,11 9,57	2,90 4,73 8,72	2,71 4,31 7,81	2,26 3,25 4,78
10	4,96 10,04 21,04	4,10 7,56 14,91	3,71 6,55 12,55	3,48 5,99 11,28	3,33 5,64 10,48	3,22 5,39 9,92	3,07 5,06 9,20	2,91 4,71 8,45	2,74 4,33 7,64	2,54 3,91 6,77	2,23 3,17 4,59
11	4,84 9,65 19,69	3,98 7,20 13,81	3,59 6,22 11,56	3,36 5,67 10,35	3,20 5,32 9,58	3,09 5,07 9,05	2,95 4,74 8,35	2,79 4,40 7,62	2,61 4,02 6,85	2,40 3,60 6,00	2,20 3,11 4,44
12	4,75 9,33 18,64	3,88 6,93 12,98	3,49 5,95 10,81	3,26 5,41 9,63	3,11 5,06 8,89	3,00 4,82 8,38	2,85 4,50 7,71	2,69 4,16 7,00	2,50 3,78 6,25	2,30 3,36 5,42	2,18 3,06 4,32
13	4,67 9,07 17,81	3,80 6,70 12,31	3,41 5,74 10,21	3,18 5.20 9,07	3,02 4,86 8,35	2,92 4,62 7,86	2,77 4,30 7,21	2,60 3,96 6,52	2,42 3,59 5,78	2,21 3,16 4,97	2,16 3,01 4,12
14	4,60 8,86 17,14	3,74 6,51 11,78	3,34 5,56 9,73	3,11 5,03 8,62	2,96 4,69 7,92	2,85 4,46 7,44	2,70 4,14 6,80	2,53 3,80 6,13	2,35 3,43 5,41	2,13 3,00 4,60	2,14 2,98 4,14
15	4,45 8,68 16,59	3,68 6,36 11,34	3,29 5,42 9,34	3,06 4,89 8,25	2,90 4,56 7,57	2,79 4,32 7,09	2,64 4,00 6,47	2,48 3,67 5,81	2,29 3,29 5,10	2,07 2,87 4,31	2,13 2,95 4,07
16	4,41 8,53 16,12	3,63 6,23 10,97	3,24 5,29 9,01	3,01 4,77 7,94	2,85 4,44 7,27	2,74 4,20 6,80	2,59 3,89 6,20	2,42 3,55 5 55	2,24 3,18 4,85	2,01 2,75 4,06	2,12 2,92 4,02

Полученные выше показатели Θ можно было бы вычислить с помощью этих двух показателей связи η и ω по формулам:

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞	t
17	4,45 8,40 15,72	3,59 6,11 10,66	3,20 5,18 8,73	2,96 4,67 7,68	2,81 4,34 7,02	2,70 4,10 6,56	2,55 3,79 5,96	2,38 3,45 5,32	2,19 3,08 4,63	1,96 2,65 3,85	2,11 2,90 3,96
18	4,41 8,28 15,38	3,55 6,01 10,39	3,16 5,09 8,49	2,93 4,58 7,46	2,77 4,25 6,81	2,06 4,01 6,35	2,51 3,71 5,76	2,34 3,37 5,13	2,15 3,01 4,45	1,92 2,57 3,67	2,10 2,88 3,92
19	4,38 8,18 15,08	3,52 5,93 10,16	3,13 5,01 8,28	2,90 4,50 7,26	2,74 4,17 6,61	2,63 3,94 6,18	2,48 3,63 5,59	2,31 3,30 4,97	2,11 2,92 4,29	1,88 2,49 3,52	2,09 2,86 3.88
20	4,35 8,10 14,82	3,49 5,85 9,95	3,10 4,94 8,10	2,87 4,43 7,10	2,71 4,10 6,46	2,60 3,87 6,02	2,45 3,56 2,44	2,28 3,23 4,82	2,08 2,86 4,15	1,84 2,42 3,38	2,09 2,84 3,85
21	4,32 8,02 14,62	3,47 5,78 9,77	3,07 4,87 7,94	2,84 4,37 6,95	2,68 4,04 6,32	2,57 3,81 5,88	2,42 3,51 5,31	2,25 3,17 4,70	2,05 2,80 4,03	1,82 2,36 3,26	2,08 2,83 3,82
22	4,30 7,94 14,38	3,44 5,72 9,61	3,05 4,82 7 80	2,82 4,31 6,81	2,66 3,99 6,19	2,55 3,75 5,76	2,40 3,45 5,19	2,23 3,12 4,58	2,03 2,75 3,92	1,78 2,30 3,15	2,07 2,82 3,79
23	4,28 7,88 14,19	3,42 5,66 9,46	3,03 4,76 7,67	2,80 4,26 6,70	2,64 3,94 6,08	2,53 3,71 5,56	2,38 3,41 5,09	2,20 3,07 4,48	2,00 2,70 3,82	1,76 2,26 3,05	2,07 2,81 3,77
24	4,26 7,82 14,03	3,40 5,61 9,34	3,01 4,72 7,55	2,78 4,22 6,59	2,62 3,90 5,98	2,51 3,67 5,55	2,36 3,36 4,99	2,18 3,03 4,39	1,98 2,66 3,74	1,73 2,21 2,97	2,06 2,80 3,75
25	4,24 7,77 13,88	3,38 5,57 9,22	2,99 4,68 7,45	2,76 4,18 6,49	2,60 3,86 5,89	2,49 3,63 5,46	2,34 3,32 4,91	2,16 2,99 4,31	1,96 2,62 3,66	1,71 2,17 2,89	2,06 2,79 3,63
26	4,22 7,72 13,74	3,37 5,53 9,12	2,98 4,64 7,36	2,74 4,14 6,41	2,59 3,82 5,80	2,47 3,59 5,38	2,32 3,29 4,83	2,15 2,96 4,24	1,95 2,58 3,59	1,69 2,13 2,82	2,06 2,78 3,61
27	4,21 7,68 13,61	3,35 5,49 9,02	2,96 4,60 7,27	2,73 4,11 6,33	2,57 3,78 5,73	2,46 3, 6 5,31	2,30 3,26 4,76	2,13 2,93 4,17	1,93 2,55 3,52	1,67 2,10 2,76	2,05 2,77 3,69
28	4,19 7,64 13,50	3,34 5,45 8,93	2,95 4,57 7,18	2,71 4,07 6,25	2,56 3,75 5,66	2,44 3,53 5,24	2,29 3,23 4,69	2,12 2,90 4,11	1,91 2,52 3,46	1,65 2,06 2,70	2,05 2,76 3,67
29	4,18 7,60 13,39	3,33 5,42 8,85	2,93 4,54 7,12	2,70 4,04 6,19	2,54 3,73 5,59	2,43 3,50 5,18	2,28 3,20 4,65	2,10 2,87 4,05	1,90 2,49 3,41	1,64 2,03 2,64	2,04 2,76 3,66
30	4,17 7,56 13,29	3,32 5,39 8,77	2,92 4,51 7,05	2,69 4,02 6,12	2,53 3,70 5,53	2,42 3,47 5,12	2,27 3,17 4,58	2,09 2,84 4,00	1,89 2,47 3,36	1,62 2,01 2,59	2,04 2,75 3,64
60	4,00 7,08 11,97	3,15 4,98 7,76	2,76 4,13 6,17	2,52 3,65 5,31	2,37 3,34 4,76	2,25 3,12 4,37	2,10 2,82 3,87	1,92 2,50 3,31	1,70 2,12 2,76	1,39 1,60 1,90	2,00 2,66 3,36
∞	3,84 6,64 10,83	2,99 4,60 6,91	2,60 3,78 5,42	2,37 3,32 4,62	2,21 3,02 4,10	2,09 2,80 3,74	1,94 2,51 3,27	1,75 2,18 2,74	1,52 1,79 2,13	1,03 1,04 1,00	1,96 2,58 3,29