

Лекция 10. Основы дисперсионного анализа

- *Общая идея дисперсионного анализа*

- *Однофакторный анализ*

- *Пример однофакторного анализа.*

1. Общая идея дисперсионного анализа

Дисперсионный анализ — это статистический метод анализа результатов наблюдений, зависящих от различных, одновременно действующих факторов, выбор наиболее важных факторов и оценка их влияния. Дисперсионный анализ находит применение в различных областях науки и техники. Идея дисперсионного анализа, как и сам термин «дисперсия», принадлежат Р. Фишеру. Суть анализа заключается в разложении общей вариации случайной величины на независимые слагаемые, каждое из которых характеризует влияние того или иного фактора или их взаимодействия.

Факторами обычно называют внешние условия, влияющие на эксперимент. Это, например, температура и атмосферное давление, сила тяготения, тип оборудования и т. п. Нас интересуют факторы, действие которых значительно и поддается проверке. В условиях эксперимента факторы могут варьировать, благодаря чему можно исследовать влияние контролируемого фактора на эксперимент. В этом случае говорят, что фактор варьирует на разных уровнях или имеет несколько уровней. В зависимости от количества факторов, включенных в анализ, различают классификацию по одному признаку — *однофакторный анализ*, по двум признакам — *двухфакторный анализ* и многостороннюю классификацию — *перекрестную классификацию*, изучением которой занимается многофакторный

анализ.

Для проведения дисперсионного анализа необходимо соблюдать следующие условия: результаты наблюдений должны быть независимыми случайными величинами, имеющими нормальное распределение и одинаковую дисперсию. Только в этом случае можно оценить значимость полученных оценок дисперсий и математических ожиданий и построить доверительные интервалы.

2. Однофакторный анализ

На практике возможен случай, когда на автоматической линии несколько станков параллельно выполняют некоторую операцию. Для правильного планирования последующей обработки важно знать, насколько однотипны средние размеры деталей, получаемые на параллельно работающих станках. Здесь имеет место лишь один фактор, влияющий на размер деталей, — станки, на которых они изготавливаются. Исследователя интересует, насколько существенно влияние этого фактора на размеры деталей? Предположим, что совокупности размеров деталей, изготовленных на каждом станке, имеют нормальное распределение и равные дисперсии. Имеем m станков, следовательно, m совокупностей или уровней, на которых произведено $n_1, n_2, \dots, n_m$ наблюдений. Для простоты рассуждений положим, что $n_1 = n_2 = \dots = n_m$. Размеры деталей, составляющие n_i наблюдений на i -м уровне, обозначим x_{i1} ,

$x_{i2}, \dots, x_{in}$. Тогда все наблюдения можно представить в виде таблицы, которую назовем *матрицей наблюдений* (табл. 10.1).

Таблица 10.1

Уровни	Наблюдения			
	1	2	j	n
1	$x_{1,1}$	$x_{1,2}$	$x_{1,j}$	$x_{1,n}$
2	$x_{2,1}$	$x_{2,2}$	$x_{2,j}$	$x_{2,n}$
...	...	...	...	...
i	$x_{i,1}$	$x_{i,2}$	$x_{i,j}$	$x_{i,n}$
...	...	...	...	...
m	$x_{m,1}$	$x_{m,2}$	$x_{m,j}$	$x_{m,n}$

Будем полагать, что для i -го уровня n наблюдений имеют среднюю β_i , равную сумме общей средней μ и вариации ее, обусловленной i -м уровнем фактора, т.е. $\beta_i = \mu + \gamma_i$. Тогда одно наблюдение можно представить в следующем виде:

$$x_{i,j} = \mu + \gamma_i + \xi_{i,j} = \beta_i + \xi_{i,j}, \quad (10.1)$$

где μ — общая средняя;

γ — эффект, обусловленный (i -м уровнем фактора;

$\xi_{i,j}$ — вариация результатов внутри отдельного уровня.

Член $\xi_{i,j}$ характеризует влияние всех не учтенных моделью (10.1) факторов. Согласно общей задаче дисперсионного анализа, нужно оценить существенность влияния фактора γ на размеры деталей. Общую вариацию переменной $x_{i,j}$ можно разложить на части, одна из которых характеризует влияние фактора γ , другая — влияние

неучтенных факторов. Для этого необходимо найти оценку общей средней μ и оценки средних по уровням β_i . Очевидно, что оценкой β является средняя арифметическая

n наблюдений 1-го уровня, т. е. $\bar{x}_{i*} = \frac{1}{n} \sum_{j=1}^n x_{i,j}$. Звездочка в

индексе при x означает, что наблюдения фиксированы на i -м уровне. Средняя арифметическая всей совокупности наблюдений является оценкой общей средней μ , т. е.

$$\bar{x} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n x_{i,j} \text{ или } \bar{x} = \frac{1}{m} \sum_{i=1}^m \bar{x}_{i*}.$$

Найдем сумму квадратов отклонений $x_{i,j}$ от x , т. е.

$\sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x})^2$. Представим ее в виде

$$\begin{aligned} Q &= \sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x})^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x}_{i*} + \bar{x}_{i*} - \bar{x})^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x}_{i*})^2 + \\ &+ \sum_{i=1}^m \sum_{j=1}^n (\bar{x}_{i*} - \bar{x})^2 + 2 \sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x}_{i*})(\bar{x}_{i*} - \bar{x}), \end{aligned} \quad (10.2)$$

обозначим

$$S = \sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x}_{i*})(\bar{x}_{i*} - \bar{x}) = \sum_{j=1}^n (x_{i,j} - \bar{x}_{i*}) \sum_{i=1}^m (\bar{x}_{i*} - \bar{x}).$$

Но $\sum_{j=1}^n (x_{i,j} - \bar{x}_{i*}) = 0$, так как это сумма отклонений

переменных одной совокупности от средней арифметической этой же совокупности, т. е. $S = 0$. Второй член суммы

(10.2) запишем в виде

$$\sum_{i=1}^m \sum_{j=1}^n (x_{i*} - \bar{x})^2 = n \sum_{j=1}^n (\bar{x}_{i*} - \bar{x})^2.$$

Тогда основное тождество (2) можно представить следующим образом:

$$\underbrace{\sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x})^2}_Q = \underbrace{n \sum_{i=1}^m (\bar{x}_{i*} - \bar{x})^2}_{Q_1} + \underbrace{\sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x}_{i*})^2}_{Q_2};$$

или $Q = Q_1 + Q_2.$ (10.3)

Слагаемое Q_1 является суммой квадратов разностей между средними уровнями и средней всей совокупности наблюдений. Эта сумма называется *суммой квадратов отклонений между группами* и характеризует расхождение между уровнями. Величину Q_1 называют также *рассеиванием по факторам*, т. е. рассеиванием за счет исследуемого фактора. Слагаемое Q_2 является суммой квадратов разностей между отдельными наблюдениями и средней i -го уровня. Эта сумма называется *суммой квадратов отклонений внутри группы* и характеризует расхождение между наблюдениями i -го уровня. Величину Q_2 называют также *остаточным рассеиванием*, т. е. рассеиванием за счет неучтенных факторов. Наконец, Q называется *общей* или *полной суммой квадратов отклонений отдельных наблюдений от общей средней $\bar{x}$* .

Зная суммы квадратов Q , Q_1 и Q_2 , можно оценить со-

ответствующие дисперсии: общую, межгрупповую и внутригрупповую. Оценим дисперсии s_1^2, s_2^2, s^2 :

$$s_1^2 = \frac{n}{m-1} \sum_{i=1}^m (\bar{x}_{i*} - \bar{x})^2 = \frac{Q_1}{m-1};$$

межгрупповая
внутригрупповая

$$s_2^2 = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_{i*})^2 = \frac{Q_2}{m(n-1)}, \quad (10.4)$$

$$s^2 = \frac{1}{mn-1} \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x})^2. \text{ общая}$$

Если влияние всех уровней фактора γ одинаково, то s_1^2 и s_2^2 — оценки общей дисперсии. Тогда для оценки существенности влияния фактора γ достаточно проверить гипотезу H_0 : $s_1^2 = s_2^2$; для этого вычисляют статистику $F = s_1^2 / s_2^2$ с $k_1 = m-1$ и $k_2 = m(n-1)$ степенями свободы. Затем по F – критерия для уровня значимости α находят критическое значение $F(\alpha, k_1, k_2)$. Если $F > F(\alpha, k_1, k_2)$, то нулевая гипотеза отвергается и делается заключение о существенном влиянии фактора γ .

При $F < F(\alpha, k_1, k_2)$ оснований отвергать нулевую гипотезу и считают, что влияние фактора γ незначительно.

Сравнивая межгрупповую и остаточную дисперсии, по

величине их отношения судят, насколько сильно проявляется влияние факторов.

Однофакторный дисперсионный анализ удобно представить в виде табл.

Таблица 10.2

Компоненты дисперсии	Сумма квадратов	Число степеней свободы k	Средний квадрат	Оценк а дисперсий
Межгрупповая	$\sum (\bar{x}_{i*} - \bar{x})^2$	$m - 1$	$\frac{n}{m-1} \sum_{i=1}^m (\bar{x}_{i*} - \bar{x})^2$	s^2_1
Внутригрупповая	$\sum_{ij} (x_{ij} - \bar{x}_{i*})^2$	$m(n - 1)$	$\frac{1}{m(n-1)} \sum_{ij} (x_{ij} - \bar{x}_{i*})^2$	s^2_2
Полная (общая)	$\sum_{ij} (x_{ij} - \bar{x})^2$	$mn - 1$	$\frac{1}{mn-1} \sum_{ij} (x_{ij} - \bar{x})^2$	s^2

3. Пример однофакторного анализа.

Пусть имеется четыре партии сырья для текстильной промышленности. Из каждой партии отобрано по пять образцов и проведены испытания на определение величины разрывной нагрузки. Результаты испытаний приведены в табл. 10.3.

Таблица 10.3

Номер партии	Разрывная нагрузка
--------------	--------------------

1	200	140	170	145	165
2	190	150	210	150	150
3	230	190	200	190	200
4	150	170	150	170	180

Требуется выяснить, существенно ли влияние различных партий сырья на величину разрывной нагрузки.

Решение. В. данном случае $m = 4$, $n = 5$. Среднюю арифметическую каждой строки вычисляем по формуле

$$\bar{x}_{i*} = \frac{1}{n} \sum_{j=1}^n x_{i,j}$$

Имеем: $\bar{x}_{1*} = 164$; $\bar{x}_{2*} = 170$; $\bar{x}_{3*} = 202$; $\bar{x}_{4*} = 164$.

Найдем среднюю арифметическую всей совокупности

$$\bar{x} = \sum_{i=1}^m \sum_{j=1}^n x_{i,j} / (mn).$$

Вычислим величины, необходимые для построения табл. 2.

сумму квадратов отклонений между группами Q_1 с $k_1 = m - 1$ степенями свободы

$$Q_1 = n \sum_{i=1}^m (\bar{x}_{i*} - \bar{x})^2 = 5 \cdot 996; k_1 = 4 - 1 = 3;$$

сумму квадратов отклонений внутри группы Q_2 с $k_1 = mn - m$ степенями свободы

$$Q_2 = \sum_{i=1}^m \sum_{j=1}^n (\bar{x}_{i,j} - \bar{x}_{i*})^2 = 7270; k_2 = 20 - 4 = 16;$$

полную сумму квадратов Q с $k = mn - 1$ степенями свободы

$$Q = \sum_{i=1}^m \sum_{j=1}^n (\bar{x}_{ij} - \bar{\bar{x}})^2 = 12250; k = 20 - 1 = 19.$$

По найденным значениям составляем табл. 10.4. Вычислим статистику: $F = (4980 \cdot 1/3) / (7270 \cdot 1/16) = 3,65$. По таблице значений F критерия при $k_2 = 16$ и $k_1 = 3$ степенях свободы и уровне значимости $\alpha = 0,01$. Имеем $F_\alpha = 9,01$. Вычисленное значение F меньше табличного, поэтому можно утверждать, что нулевая гипотеза не отвергается, а это значит, что различие между сырьем в партиях не влияет на величину разрывной нагрузки.

Таблица 10.4

Компонент дисперсии	Суммы квадратов	Число степеней	Средний квадрат
Межгрупповая	4 980	3	1660,0
Внутригрупповая	7 270	1	454,4
Полная	12 250	6	644,7

Следует осторожно подходить к истолкованию окончательных результатов, так как они предполагают нормальную плотность и тождественность дисперсий. Каждое из допущений требует проверки, основанной на тщательном анализе проведенных экспериментов.

Лекция 11. Двухфакторный дисперсионный анализ

- задача двухфакторного ДА;
- общий случай двухфакторного ДА
- пример двухфакторного ДА

1. Задача двухфакторного дисперсионного анализа.

Если на результативный признак влияют несколько факторов одновременно, то имеет место многофакторный анализ. Дисперсионный анализ в этом случае имеет свои особенности, так как необходимо учитывать взаимодействия между факторами.

Рассмотрим задачу оценки влияния двух одновременно действующих факторов. Предположим, что имеется несколько однотипных станков и несколько видов сырья. Требуется выяснить, значимо ли влияние различных станков и качества сырья в партиях на качество обрабатываемых деталей. Это типичная задача двухфакторного дисперсионного анализа.

Считаем, что предпосылки дисперсионного анализа выполнены.

Пусть:

фактор А — влияние настройки станка,

фактор В — влияние качества сырья.

Имеем r станков, следовательно, r уровней фактора А, v партий сырья, следовательно, v уровней фактора В. Матрицу наблюдений можно представить в виде табл. 11.5. Пересечение 1-го уровня фактора F с i -м уровнем фактора В образует ij - ю ячейку, в которую записывают наблюдения, полученные при одновременном исследовании факторов А и В на i -м и j -м уровнях.

Для простоты можно предположить, что имеем в ячейке только одно наблюдение x_{ij} . Предположим также, что между факторами А и В нет взаимодействия и что на i -м уровне фактора А наблюдения имеют среднюю β_{iA} , а на j -том уровне фактора В наблюдения – среднюю β_{jB} .

Таблица 11.5

Партии сырья j Ставки i	B_1	B_2	...	B_j	...	B_v	$\tilde{X}_{i*}$
A_1	x_{11}	x_{12}	...	x_{1j}	...	x_{1v}	$\tilde{X}_{i*}$
A_2	x_{11}	x_{22}	...	x_{1j}	...	x_{1v}	$\tilde{X}_{i*}$
...	...	...	...	...	...	..	...
A_i	x_{11}	x_{i2}	...	x_{1j}	...	x_{1v}	$\tilde{X}_{i*}$
...	...	...	...	...	...	...	...
A_r	x_{11}	x_{r2}	...	x_{1j}	...	x_{1v}	$\tilde{X}_{i*}$
$\tilde{X}_{*j}$	$\tilde{X}_{*1}$	$\tilde{X}_{*2}$	...	$\tilde{X}_{*j}$	...	$\tilde{X}_{*v}$	$\bar{X}$

Тогда одно наблюдение можно представить в виде

$$x_{i,j} = \mu + \beta_i + \lambda_j + \delta_{ij} \quad (11.5)$$

где μ - общая средняя; β - эффект, обусловленный влиянием i - го уровня фактора А λ - эффект, обусловленный влиянием j - го уровня фактора В; $\delta_{i,j}$ вариация результатов внутри отдельной ячейки (в случае одного наблюдения вариация равна нулю).

Оценками μ , β_{iA} и λ_{jB} являются соответственно общая средняя

$$\bar{X} = \frac{1}{r \cdot v} \sum_{i=1}^r \sum_{j=1}^v x_{i,j}$$

и средние по уровням

$$\bar{x}_{i*} = \frac{1}{v} \sum_{j=1}^v x_{i,j} \quad \bar{x}_{*j} = \frac{1}{r} \sum_{i=1}^r x_{i,j}$$

Оценки общей дисперсии можно получить из основного тождества дисперсионного анализа. В двухфакторном дисперсионном анализе общая сумма квадратов отклонений от общей средней раскладывается согласно формуле (4) уже не на две, а на три части: часть общей суммы квадратов, обусловленную влиянием фактора A , часть, обусловленную влиянием фактора B , и часть, обусловленную влиянием неучтенных факторов. С помощью дисперсионных отношений можно выяснить, насколько существенно влияние каждой из этих частей. Действительно,

$$\begin{aligned} Q &= \sum_{i=1}^r \sum_{j=1}^v (x_{ij} - \bar{x})^2 = \sum_{i=1}^r \sum_{j=1}^v (x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x} + \bar{x}_{i*} - \bar{x} + \bar{x}_{*j} - \bar{x})^2 = \\ &= v \sum_{i=1}^r (\bar{x}_{i*} - \bar{x})^2 + r \sum_{j=1}^v (\bar{x}_{*j} - \bar{x})^2 + \sum_{i=1}^r \sum_{j=1}^v (x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2 = Q_1 + Q_2 + Q_3; \end{aligned}$$

Слагаемое Q_1 представляет собой сумму квадратов разностей между средними по строкам и общим средним и характеризует изменение признака по фактору A . Слагаемое Q_2 представляет собой сумму квадратов разностей между средними по столбцам и общим средним и характеризует изменение признака по фактору B . Слагаемое Q_3 называется остаточной суммой квадратов и характеризует влияние неучтенных факторов. Сумма Q называется общей или полной суммой квадратов отклонений отдельных наблюдений от общей средней. Оценим дисперсии;

$$s^2 = \frac{1}{rv-1} \sum_{i=1}^r \sum_{j=1}^v (x_{ij} - \bar{x})^2 = \frac{Q}{rv-1}, \quad (11.6)$$

$$s_1^2 = \frac{v}{r-1} \sum_{i=1}^r (\bar{x}_{i*} - \bar{x})^2 = \frac{Q_1}{r-1}, \quad (11.7)$$

$$s_2^2 = \frac{r}{v-1} \sum_{j=1}^v (\bar{x}_{*j} - \bar{x})^2 = \frac{Q_2}{v-1}, \quad (11.8)$$

$$s_3^2 = \frac{1}{(r-1) \cdot (v-1)} \sum_{i=1}^r \sum_{j=1}^v (x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2 = \frac{Q_3}{(r-1) \cdot (v-1)}. \quad (11.9)$$

В двухфакторном анализе для выяснения значимости влияния факторов А и В на исследуемый признак сравнивают дисперсии по факторам с остаточной дисперсией. Вычисляют статистики $F_1 = s_1^2/s_2^2$ с $(r-1)$ и $(r-1)(v-1)$ степенями свободы и $F_2 = s_2^2/s_3^2$ с $(v-1)$ и $(r-1)(v-1)$ степенями свободы. Сравнение вычисленных статистик с табличными значениями и выводы о существенности влияния факторов производят так же, как и в однофакторном дисперсионном анализе.

Двухфакторный дисперсионный анализ удобно представить в виде табл. 11.6

Таблица 11.6

Компонента дисперсии	Сумма квадратов	Число степеней	Оцен
----------------------	-----------------	----------------	------

		свободы	ки дисп ерсий
Между средними по строкам	$Q_1 = v \sum_{i=1}^r (\bar{x}_{i*} - \bar{x})^2$	$r - 1$	s_1^2
Между средними по столбцам	$Q_2 = r \sum_{j=1}^v (\bar{x}_{*j} - \bar{x})^2$	$v - 1$	s_2^2
Остаточная	$Q_3 = \sum_{i=1}^r \sum_{j=1}^v (\bar{x}_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2$	$(r - 1) * (v - 1)$	s_3^2
Полная (общая)	$Q = \sum_{i=1}^r \sum_{j=1}^v (\bar{x}_{ij} - \bar{x})^2$	$rv - 1$	s^2

Рассмотрим на примере построение двухфакторного комплекса с одним наблюдением в ячейке. Имеем три уровня фактора В: B_1 , B_2 и B_3 и два уровня фактора А:

A_1 и A_2 (табл. 7). Для данного комплекса $r = 2$, $v = 3$, $n = rv = 6$. В нижней строке табл. 7 и в правом крайнем столбце приведены средние значения по строкам и столбцам, т. е. по уровням факторов. Так, среднее по уровню фактора B_1 равно

$$\bar{x}_{*1} = (1 + 5) / 2 = 3;$$

Т а б л и ц а 11.7

A \ B				
	B_1	B_2	B_3	$\bar{x}_{i*}$

A_1	1	2	3	2
A_2	5	6	10	7
$\bar{X}_{*j}$	3	4	6,5	4,5

среднее по уровню фактора A_1 равно $(1+2+3)/3 = 2$. Общее среднее $\bar{x} = 4,5$. По формуле (5) вычислим суммы квадратов:

$$Q_1 = 37,5; Q_2 = 13; Q_3 = 3; Q = 53,5.$$

Для нахождения оценок дисперсий воспользуемся формулами (11.6 – 11.9). Имеем:

$$s_1^2 = 37,5/1 = 37,5; \quad s_2^2 = 13/2 = 6,5; \quad s_3^2 = 3/1 \cdot 2 = 1,5.$$

Теперь табл. 11.6 можно записать в следующем виде (табл. 7.8).

Таблица 11.8

Компонента дисперсий	Сумма квадратов	Число степеней свободы	Оценка дисперсии
Между средними по строкам (фактор A)	37,5	1	37,5
Между средними по столбцам (фактор B)	13,0	2	6,5
Остаточная	3,0	2	1,5
Полная	53,5	5	10,7

Вычисляем F_A и F_B :

$$F_A = s_1^2/s_3^2 = 37,5/1,5 = 25; F_B = s_2^2/s_3^2 = 6,5/1,5 = 4,3.$$

Для уровня значимости $\alpha = 0,05$ и $k_2 = 2$, $k_1 = 1$ степеней свободы по табл. 4 приложений находим значения F_α :

$$F_{\alpha(A)} = 18,51. F_A = 25; \quad F_{\alpha(B)} = 19,0; \quad F_B = 4,3.$$

Сравнивая табличные значения с вычисленными, имеем:

$$F_A > F_{\alpha(A)} \quad F_B < F_{\alpha(B)}.$$

Полученные результаты позволяют сделать следующие выводы:

1. Нулевая гипотеза о равенстве средних по строкам не подтверждается, т. е. влияние фактора А на исследуемый признак значимо.
2. Нулевая гипотеза о равенстве средних по столбцам не опровергается, т. е. влияние фактора В на исследуемый признак незначимо.

Мы рассмотрели частный случай дисперсионного анализа при классификации по двум признакам: в ячейке одно наблюдение, взаимодействие между факторами отсутствует. В общем случае в ячейке может быть несколько наблюдений (как равное количество, так и неравное), между факторами может иметь место взаимодействие. Лучше, когда в ячейке равное количество наблюдений, так как при этом упрощаются вычисления.

2. Общий случай двухфакторного ДА

Для общего случая двухфакторного анализа одно наблюдение можно представить в виде

$$x_{ijk} = \mu + \gamma_i + g_j + v_{ij} + e_{ijk}$$

где μ - общее среднее;

γ_i - эффект, обусловленный влиянием i - го уровня фактора А;

g_j - эффект, обусловленный влиянием j - го уровня фактора В;

v_{ij} - эффект взаимодействия факторов А и В;

e_{ijk} - вариация внутри ячейки.

Основное тождество двухфакторного дисперсионного

анализа с одинаковым количеством наблюдений в ячейке (n) имеет вид

$$Q = \sum_{i=1}^r \sum_{j=1}^v \sum_{k=1}^n (x_{ijk} - \bar{x})^2 = nv \sum_{i=1}^r (\bar{x}_{i**} - \bar{x})^2 + nr \sum_{j=1}^v (\bar{x}_{*j*} - \bar{x})^2 + n \sum_{i=1}^r \sum_{j=1}^v (\bar{x}_{ij*} - \bar{x}_{i**} - \bar{x}_{*j*} + \bar{x})^2 + \sum_{i=1}^r \sum_{j=1}^v \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij*})^2 = Q_1 + Q_2 + Q_3 + Q_4.$$

Здесь Q_1 и Q_2 имеют то же значение, что и в формуле (7.5);

Q_3 — сумма квадратов, оценивающая взаимодействие факторов А и В;

Q_4 — сумма квадратов, оценивающая вариацию внутри ячейки.

Матрицу наблюдений можно представить в виде табл. 11.9.

Таблица 11.9

A \ B					
	B ₁	B ₂	...	B _v	$\bar{x}_{i**}$
A ₁	$\bar{x}_{11*}$	$\bar{x}_{12*}$	...	$\bar{x}_{1v*}$	$\bar{x}_{1**}$
A ₂	$\bar{x}_{21*}$	$\bar{x}_{22*}$	...	$\bar{x}_{2v*}$	$\bar{x}_{2**}$
...	...		...	...	...
A _r	$\bar{x}_{r1*}$	$\bar{x}_{r2*}$	...	$\bar{x}_{rv*}$	$\bar{x}_{r**}$
$\bar{x}_{*j*}$	$\bar{x}_{*1*}$	$\bar{x}_{*2*}$	...	$\bar{x}_{*v*}$	$\bar{x}$

Здесь $x_{111}, x_{112}, \dots, x_{rvn}$ - наблюдавшиеся значения исследуемого признака;

$$\bar{x}_{ij*} = \frac{1}{n} \sum_{k=1}^n x_{ijk} \quad - \text{среднее значение в ячейке};$$

$$\bar{x}_{i**} = \frac{1}{v} \sum_{k=1}^v x_{ij*} \quad - \text{среднее значение по строке};$$

$$\bar{x}_{*j*} = \frac{1}{r} \sum_{i=1}^r x_{ij*} \quad - \text{среднее значение по столбцу};$$

$$\bar{x}_{**} = \frac{1}{rv} \sum_{i=1}^r \sum_{j=1}^v x_{ij*} \quad - \text{общее среднее}; \quad (11.10)$$

r - число уровней фактора А;

v - число уровней фактора В.

Порядок проведения дисперсионного анализа в этом случае такой же, как и прежде: сначала вычисляют суммы квадратов, оценки дисперсий, затем отношение дисперсий F сравнивают с табличным.

Схема анализа и порядок вычисления сумм приведены в табл. 11.10

Таблица 11.10

Компонента дисперсий	Суммы квадратов	Число степеней свободы	Оценка дисперсии
----------------------	-----------------	------------------------	------------------

Между средними по строкам (по фактору А)	$Q_1 = vn \sum_{i=1}^r (x_{i**} - \bar{x})^2$	$v - 1$	s_1^2
Между средними по столбцам (по фактору В)	$Q_2 = rn \sum_{j=1}^v (x_{*j*} - \bar{x})^2$	$r - 1$	s_2^2
Взаимодействие		$(v - 1)^*$	s_3^2
Остаточная	$Q_3 = n \sum_{i=1}^r \sum_{j=1}^v (x_{ijk} - \bar{x}_{i**} - \bar{x}_{*j*} + \bar{x})^2$	$(r - 1)$	s_4^2
Полная сумма квадратов	$Q_4 = \sum_{i=1}^r \sum_{j=1}^v \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij*})^2$ $Q = \sum_{i=1}^r \sum_{j=1}^v \sum_{k=1}^n (x_{ijk} - \bar{x})^2$	$rv(n - 1)$ $rvn - 1$	s^2

Как видно из табл. 11.10, в схеме анализа появляется новая сумма квадратов Q_4 и несколько меняется структура суммы Q_3 (вместо x_{ijk} берется x_{ij*}). Появление суммы Q_4 обусловлено наличием нескольких наблюдений в ячейке. В предыдущей схеме (см. табл. 6) эта сумма отсутствовала, так как при одном наблюдении в ячейке разность $(x_{ijk} - \bar{x}_{ij*})$ равна нулю. Сумма Q_4 характеризует влияние прочих случайных факторов (кроме факторов А, В и их взаимодействия), поэтому для определения значимости влияния факторов А и В величину дисперсии, обусловленную влиянием этих факторов, сравнивают с дисперсией, обусловленной влиянием прочих факторов. При этом вычисляют следующие отношения дисперсий:

$$F_A = s_1^2/s_4^2; \quad F_B = s_2^2/s_4^2; \quad F_{AB} = s_3^2/s_4^2;$$

Вычисленные значения сравнивают с табличными значениями F_α , которые получены для заданного уровня значимости и соответствующего числа степеней свободы.

3. Пример двухфакторного ДА

Рассмотрим пример построения двухфакторного комплекса по приведенной схеме. В текстильной промышленности важным является выявление факторов, влияющих на качество пряжи, с тем чтобы в дальнейшем их было можно регулировать. В табл. 11.11 приведены данные о величине разрывной нагрузки в зависимости от наладки машины и партии сырья.

Таблица 11.11

Партии сырья Уровень наладки станка	В ₁					В ₂				
	1	2	3	4	5	1	2	3	4	5
A ₁	190	260	170	170	170	190	150	210	150	150
A ₂	150	250	220	140	180	230	190	200	190	200
A ₃	190	185	135	195	195	150	170	150	170	180

При каждом уровне А наладки машины исследовано по пять образцов из каждой партии В сырья для определения разрывной нагрузки. Требуется выяснить, значимо ли влияют наладка машины и партии сырья на величину разрывной нагрузки.

По каждому уровню факторов А и В определяем средние значения. Табл. 11.11 преобразуется к следующему виду (табл. 11.12).

Таблица 11.12

Партии сырья Уровень наладки машины	B_1	B_2	$\bar{x}_{i**}$
A_1	$\bar{x}_{11*} = 192$	$\bar{x}_{12*} = 170$	$\bar{x}_{1**} = 181$
A_2	$\bar{x}_{21*} = 188$	$\bar{x}_{22*} = 202$	$\bar{x}_{2**} = 195$
A_3	$\bar{x}_{31*} = 180$	$\bar{x}_{32*} = 164$	$\bar{x}_{3**} = 172$
$\bar{x}_{*j*}$	$\bar{x}_{*1*} = 186,7$	$\bar{x}_{*2*} = 178,7$	$\bar{x} = 182,7$

Находим суммы квадратов (см. табл. 11.10)

$Q_1 = 2686,6$; $Q_2 = 480$; $Q_3 = 1860$; $Q_4 = 2260$; $Q = Q_1 + Q_2 + Q_3 + Q_4 = 27386,7$.

Вычисляем оценки дисперсий:

$s_1^2 = 2686,7$; $s_2^2 = 240$; $s_3^2 = 930$; $s^2 = 944,4$; $s_4^2 = 931,7$.

Табл. 10 в нашем случае преобразуется к следующему виду (табл. 11.13).

Таблица 11.13

Компонента дисперсии	Суммы квадратов	Число степе- ней свободы	Оценка дисперсии
-------------------------	--------------------	-----------------------------	---------------------

Между средними по строкам (по фактору А)	2686,7 480	3-1=2 2-1=1	2686,7 240 .
Между средними по столбцам (по фактору В)	1860	2·1=2	930
Взаимодействие			
Остаточная	22360	3·2·(5-1)=24	931,7
Полная	27386,7	3·2·5=30	944,4

Вычисляем отношения дисперсий: $F_A = 2686,7 / 931,7 = 2,88$; $F_B = 240 / 931,7 = 0,26$.

При уровне значимости $\alpha = 0,05$; $k_4 = 24$ и $k_1 = 1$ степенями свободы для F_A и $k_2 = 24$, $k_2 = 2$ степеням свободы для F_B находим следующие табличные значения: $F_{\alpha(A)} = 4,26$; $F_{\alpha(B)} = 3,40$.

Сравнивая табличные значения с вычисленными, имеем $F_A < F_{\alpha(A)}$ и $F_B < F_{\alpha(B)}$. Следовательно, нулевая гипотеза о равенстве средних не отвергается, т. е. влияние фактора А (уровня наладки машины) и фактора В (партии сырья) на величину разрывной нагрузки незначимо.

Критические значения функции χ^2 - распределения

p			n
0,05	0,025	0,01	
3,84	5,02	6,63	1
5,99	7,38	9,21	2
7,81	9,35	11,3	3
9,49	11,1	13,3	4
11,1	12,8	15,1	5
15,5	17,5	20,1	8
18,3	20,5	23,2	10

25,0	27,5	30,6	15
31,4	34,2	37,6	20
43,8	47,0'	50,9	30
55,8	59,3	63,7	40
67,5	71,4	76,2	50

Критические значения функции распределения Фишера $F_p(f_1, f_2)$

f_2	f_1										
	1	2	3	5	7	9	10	15	20	40	∞
p = 0,01											
1	4052,2	4999,5	5403,3	5763,7	5928,3	6022,5	6055,8	6157,3	6208,7	6286,8	6366,0
2	98,50	99,00	99,17	99,29	99,36	99,38	99,40	99,43	99,45	99,47	99,50
3	34,12	30,82	29,46	28,28	27,65	27,35	27,23	26,87	26,69	26,41	26,13
5	16,26	13,27	12,06	10,97	10,46	10,16	10,05	9,72	9,55	2,29	9,02
7	12,27	9,55	8,45	7,46	6,99	6,72	6,62	6,43	6,16	5,91	5,65
9	10,56	8,02	6,99	6,06	5,61	5,35	5,26	4,96	4,81	4,57	4,31
10	10,04	7,56	6,55	5,64	5,20	4,94	4,85	4,56	4,41	4,17	3,91
15	8,68	6,36	5,42	4,56	4,14	3,90	3,81	3,52	3,37	3,13	2,87
20	8,10	5,85	4,94	4,10	3,70	3,46	3,37	3,09	2,94	2,67	2,42
40	7,31	5,18	4,31	3,51	3,12	2,89	2,80	2,52	2,37	2,11	1,80
∞	6,64	4,61	3,78	3,02	2,64	2,40	2,32	2,04	1,88	1,59	1,00
p = 0,05											
1	161,5	199,5	215,7	230,1	236,8	240,5	241,88	249,95	248,01	251,14	254,32
2	18,51	19,00	19,16	19,29	19,35	19,38	19,40	19,43	19,45	19,47	19,50
3	10,13	9,55	9,28	9,01	8,89	8,81	5,79	8,70	8,66	8,59	8,53
5	6,60	5,77	5,41	5,05	4,88	4,77	4,74	4,62	4,56	4,46	4,37
7	5,59	4,74	4,35	3,97	3,78	3,68	3,64	3,51	3,45	3,34	3,23
9	5,12	4,26	3,86	3,48	3,29	3,18	3,14	3,01	2,94	2,83	2,71
10	4,97	4,10	3,71	3,33	3,14	3,02	2,98	2,85	2,77	2,66	2,54
15	4,54	3,68	3,29	2,90	2,70	2,56	2,54	2,40	2,33	2,20	2,07
20	4,35	3,49	3,10	2,71	2,51	2,39	2,35	2,20	2,12	1,99	1,84
40	4,09	3,23	2,83	2,45	2,25	2,12	2,08	1,93	1,84	1,69	1,51
∞	3,84	2,99	2,60	2,21	2,01	1,88	1,83	1,67	1,57	1,39	1,00

Варианты заданий (номер задания = номер в журнале делить на 2)

тема «Одно и двухфакторный дисперсионный анализ»

1. Для однофакторного анализа столбцы это наблюдения, а строки уровни.
2. Для двух факторного первые 5 столбцов – первый фактор, остальные – второй фактор

1

	0	1	2	3	4	5	6	7	8	9
0	207	249	148	181	186	173	129	228	131	165
1	154	251	229	131	176	225	191	208	182	216
2	166	198	106	181	192	166	156	129	173	173

2

	0	1	2	3	4	5	6	7	8	9
0	178	257	159	199	187	178	159	231	161	153
1	140	245	196	134	191	210	186	227	182	209
2	187	185	141	193	211	157	171	155	175	194

3

	0	1	2	3	4	5	6	7	8	9
0	186	252	181	178	149	197	159	214	162	149
1	139	299	199	138	181	237	192	186	188	163
2	191	181	155	193	173	168	191	147	170	203

4

	0	1	2	3	4	5	6	7	8	9
0	196	258	154	176	176	173	128	202	151	150
1	156	249	230	154	204	246	184	229	200	182
2	173	193	145	194	194	149	156	129	175	167

5

	0	1	2	3	4	5	6	7	8	9
0	196	288	168	165	170	220	111	242	140	140
1	166	255	231	151	169	222	182	193	160	210
2	200	187	146	198	210	165	178	157	184	196

6

	0	1	2	3	4	5	6	7	8	9
0	203	242	191	172	166	177	124	240	112	132
1	125	266	247	141	195	237	184	199	201	208
2	182	188	130	210	216	117	147	155	185	221

7

	0	1	2	3	4	5	6	7	8	9
0	177	265	160	170	188	176	163	240	160	167
1	166	248	242	124	182	212	193	227	194	223
2	187	193	139	197	222	146	166	129	176	186

8

	0	1	2	3	4	5	6	7	8	9
0	186	257	149	162	175	178	133	220	145	152
1	165	240	214	122	182	187	178	186	187	218
2	187	174	140	222	208	156	176	145	159	178

9

	0	1	2	3	4	5	6	7	8	9
0	193	282	155	162	188	208	160	230	165	134
1	126	253	224	165	156	239	162	201	192	194
2	203	179	115	183	202	169	167	161	162	164

10

	0	1	2	3	4	5	6	7	8	9
0	212	254	172	159	150	184	145	181	152	145
1	133	262	229	128	190	264	198	211	196	227
2	197	185	147	191	182	135	148	141	158	189

11

	0	1	2	3	4	5	6	7	8	9
0	179	256	182	141	177	189	152	225	165	132
1	147	252	233	146	182	226	199	184	190	211
2	208	167	114	201	211	127	188	146	167	196

12

	0	1	2	3	4	5	6	7	8	9
0	193	259	186	176	155	220	161	203	150	158
1	153	228	229	143	191	240	212	179	201	215
2	198	162	140	188	194	119	173	164	164	187

13

	0	1	2	3	4	5	6	7	8	9
0	186	249	171	161	160	175	152	215	147	155
1	155	258	219	131	172	207	182	213	171	193
2	204	192	147	179	194	145	190	172	151	173

14

	0	1	2	3	4	5	6	7	8	9
0	193	264	134	151	190	183	163	241	158	140
1	158	263	204	156	185	225	181	203	201	183
2	211	180	123	179	191	146	174	157	170	152

15

	0	1	2	3	4	5	6	7	8	9
0	201	250	172	197	176	179	171	196	155	151
1	162	259	193	139	191	264	191	209	174	201
2	193	198	122	200	200	165	148	142	176	194

16

	0	1	2	3	4	5	6	7	8	9
0	177	258	149	153	156	213	135	204	169	137
1	169	237	237	123	194	233	190	192	173	186
2	204	168	144	231	200	145	170	151	174	183