Дисперсионный анализ

Однофакторный дисперсионный анализ

Дисперсионный анализ

- Данный вид анализа применяют в тех случаях, когда необходимо сопоставить не 2, а большее число результатов однотипных экспериментов.
- Смысл дисперсионного анализа заключается в следующем – из общей суммы квадратов дисперсии вычитают сумму квадратов отклонений по изучаемым факторам (межфакторная дисперсия). В результате чего получают остаточную сумму квадратов дисперсии, которая характеризует влияние различных факторов.

Таблица исходных данных

Уровень факторов	Результаты						$\sum$	Число наблюдений	Среднее арифметическое
	1	2	......	j	......	m			
1	y_{11}	y_{12}		y_{1j}		y_{1m}	$\sum_{j=1}^{m} y_{1j}$	m_1	$\bar{y}_1$
2									
......									
i	y_{i1}	y_{i2}		y_{ij}		y_{im}	$\sum_{j=1}^{m} y_{ij}$	m_i	$\bar{y}_i$
....									
n	y_{n1}	y_{n2}		y_{nj}		y_{nm}	$\sum_{j=1}^{m} y_{nj}$	m_n	$\bar{y}_m$
Итого							$\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}$	$N=m\cdot n$	$\bar{y}$

Для n разных уровней некоторого фактора проводят по m измерений (для каждого уровня) величины y. Затем проверяют гипотезу о том, что влияние фактора на средние значения для каждого уровня существенно.

Расчетные формулы

- Вычисляют общее среднее и среднее для данного уровня фактора. $\bar{y}$ $\bar{y}_i$
- Вычисляют остаточную дисперсию:

$$S_r^2 = \left[\sum_{i=1}^{n} \sum_{j=1}^{m} \left(y_{ij} - y_i \right)^2 \right] \Big/ (mn - m)$$

- Вычисляют межфакторную дисперсию:
-

$$S_f^2 = m \left[\sum_{i=1}^{n} \left(\bar{y}_i - \bar{y} \right)^2 \right] \Big/ (n - 1)$$

Дисперсия генеральной совокупности:

$$S_0^2 = \left[\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-y_i)^2\right] \Big/ (mn-1)$$

Определяют статистику:

$$F^* = S_f^2 / S_r^2$$

Сравнивают ее с табличным значением $F_\alpha(v_1, v_2)$, $((v_1=n-1), (v_2=nm-n))$.

$F^* > F_\alpha(v_1, v_2)$, фактор статистически значим

Дисперсионный анализ

Двухфакторный анализ

Исходные данные

- Для двух факторного анализа необходимо задать **m**, как число измерений величины **y**.
- Пусть второй фактор **k** – принимает значения от 1 до **p**, а первый фактор – от 1 до **n**.
- Запишем **y** с тремя индексами **k**, **i**, **j**, где **j** – число повторений измерения (y_{kij}).

Проверяются три гипотезы:

- Влияние первого фактора – статистически значимое.
- Влияние второго фактора – статистически значимое.
- Взаимодействие между факторами – статистически значимое.

Вычисления ведут по следующей схеме:

- Определяют общее и частное среднее.

$$\bar{y} = \frac{\sum_{k=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{kij}}{mnp} \qquad \bar{y}_{ki} = \frac{\sum_{j=1}^{m} y_{kij}}{m} \qquad \bar{y}_k = \frac{\sum_{i=1}^{n} \bar{y}_{ki}}{n} \qquad \bar{y}_i = \frac{\sum_{k=1}^{p} \bar{y}_{ki}}{p}$$

- Дисперсия генеральной совокупности

$$S_0^2 = \frac{\sum_{k=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m} \left(y_{kij} - \bar{y}\right)^2}{mnp-1}$$

- Межфакторные S1 и S2:

$$S_1^2 = \frac{mp\left[\sum_{i=1}^{n} (\bar{y}_i - \bar{y})^2\right]}{n-1} \qquad S_2^2 = \frac{mn\left[\sum_{k=1}^{p} (\bar{y}_k - \bar{y})^2\right]}{p-1}$$

- Дисперсия взаимодействия:

$$S_{12} = \frac{m\left[\sum_{i=1}^{n}\sum_{k=1}^{p}\left(\bar{y}_{ki} - \bar{y}_{i} - \bar{y}_{k} + \bar{y}\right)^{2}\right]}{(n-1)\cdot(p-1)}$$

- Остаточная дисперсия:

$$S_{r}^{2} = \frac{\sum_{k=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{kij} - \bar{y}_{ki}\right)^{2}}{pn(m-1)}$$

- Проверка статистической значимости:

$$F_{1} = \frac{S_{1}^{2}}{S_{r}^{2}} \qquad F_{табл} = \left(n-1;\, pn(m-1)\right)$$

$$F_{2} = \frac{S_{2}^{2}}{S_{r}^{2}} \qquad F_{табл} = \left(p-1;\, pn(m-1)\right)$$

$$F_{3} = \frac{S_{12}^{2}}{S_{r}^{2}} \qquad F_{табл} = \left((n-1)\cdot(p-1);\, pn(m-1)\right)$$

Пример

- необходимо проанализировать зависимость пластичности полуфабриката от размера субзерна (1 фактор) и объемной доли избыточных фаз (2 фактор). В каждой точке испытывали по три образца

		Первый фактор			
№ \ №		1	2	3	4
Второй фактор	1	2,5 ; 2,8 ; 2,6	2,6 ; 2,6 ; 2,7	3,0 ; 3,2 ; 3,4	2,9 ; 2,4 ; 2,3
	2	3,5 ; 3,8 ; 3,6	3,3 ; 3,8 ; 3,6	3,7 ; 3,7 ; 3,3	3,9 ; 3,8 ; 3,9
	3	4,5 ; 4,9 ; 4,7	4,2 ; 4,9 ; 4,9	4,6 ; 4,4 ; 4,4	4,1 ; 4,8 ; 4,9
	4	6,0 ; 6,3 ; 6,6	6,3 ; 6,3 ; 6,5	6,0 ; 6,0 ; 7,0	6,0 ; 6,3 ; 6,6
	5	8,0 ; 8,3 ; 8,5	8,6 ; 8,0 ; 8,1	8,0 ; 8,5 ; 8,2	8,5 ; 8,5 ; 8,1

Расчеты

- $S_0^2 =$ 1,41; $S_1^2 =$ 0,081 ; $S_2^2 =$ 632,9 ; $S_{12}^2 =$ 0,231 ; $S_r^2 =$ 0,095.
- Поделив друг на друга, сравниваем с табличными значениями:
- Для 1-го фактора: $F_{0.05}(3/40)=2,84$
- Для 2-го фактора: $F_{0,05}(4/40)=2,61$
- Для взаимодействия: $F_{0,05}(12/40)=2,00$
- После расчета выясняем, что размер субзерна не влияет на пластичность полуфабриката, на него влияет лишь доля избыточных фаз.

Дисперсионный анализ

Латинские квадраты

Латинский квадрат

- – это квадратная таблица размером **n** x **n** элементов, расположенных на поле квадрата таким образом, что каждый из них встречается в каждом столбце и в каждой строке только по одному разу.

	1	2	3	4
1	*a*	*b*	*c*	*d*
2	*b*	*c*	*d*	*a*
3	*c*	*d*	*a*	*b*
4	*d*	*a*	*b*	*c*

- Если строкам, столбцам и элементам выписной таблицы поставить в соответствие уровни каких – либо факторов (А, В, С), то латинский квадрат можно рассматривать как план из эксперимента, позволяющий провести дисперсионный анализ с тремя факторами.
-

Оценка статистической значимости

$$y = \mu + A_i + B_j + C_k + E_{ijk}$$

E_{ijk} – эффект взаимодействия.

ПФЭ=4^3=

$$y^3 = 64 опыта <=> 4 \times 4 = 16 - дисперсионный\ \ анализ$$

Латинский квадрат 4x4

Уровень фактора B	Уровень фактора A			
	A_1	A_2	A_3	A_4
B_1	C_1	C_2	C_3	C_4
B_2	C_2	C_3	C_4	C_1
B_3	C_3	C_4	C_1	C_2
B_4	C_4	C_1	C_4	C_3

При проведении эксперимента порядок реализации опытов необходимо рандомизировать, статистические свойства оценок при этом улучшаются.

Ортогональные планы

- Латинские квадраты называются взаимно ортогональными, если при размещении их элементов на поле общей таблицы каждая пара элементов двух квадратов встречается только по одному разу.

$$
\begin{array}{ccc} a & b & c \\ b & c & a \\ c & a & b \end{array}
\;\star\;
\begin{array}{ccc} \alpha & \beta & \gamma \\ \gamma & \alpha & \beta \\ \beta & \gamma & \alpha \end{array}
\;\Rightarrow\;
\begin{array}{ccc} a\alpha & b\beta & c\gamma \\ b\gamma & c\alpha & a\beta \\ c\beta & a\gamma & b\alpha \end{array}
$$

Матрица плана эксперимента

- На базе латинского квадрата 3х3

Уровень B	Уровень A		
	A_1	A_2	A_3
B_1 B_2 B_3	C_1D_1 C_2D_3 C_3D_2	C_2D_2 C_3D_1 C_1D_2	C_3D_3 C_1D_2 C_2D_1

Уровни факторов C и D располагаются по полю таблицы в виде ортогонального плана

Методика расчета

- Вычисление сумм результатов по строкам, столбцам и одноименным буквам: ∑ A,B,C,D (по отдельности).
- Вычисление вспомогательных расчетных сумм:

$$SS_1 = \sum_{p=1}^{n^2} \left(y_p\right)^2$$

- где p – текущий индекс ячейки квадратов

Средние суммы квадратов по строкам, столбцам и латинским буквам:

$$SS_2 = \frac{\sum_{i=1}^{n} A_i^2}{n} \qquad SS_3 = \frac{\sum_{j=1}^{n} B_j^2}{n} \qquad SS_4 = \frac{\sum_{k=1}^{n} C_k^2}{n} \qquad SS_5 = \frac{\sum_{l=1}^{n} D_l^2}{n}$$

;

; Вычисление корректирующего члена

$$SS_6 = \frac{\left(\sum_{p=1}^{n^2} y_p\right)^2}{n^2}$$

; Вычисление сумм квадратов, характеризующих эффекты строк,
; столбцов, латинских букв

$$SS_A = SS_2 - SS_6 \qquad SS_C = SS_4 - SS_6$$

$$SS_B = SS_3 - SS_6 \qquad SS_D = SS_5 - SS_6$$

Вычисление остаточной суммы квадратов

$$SS_{ост.} = SS_1 - \left(SS_A + SS_B + SS_C + SS_D + SS_6\right)$$

Вычисление оценок эффектов строк, столбцов и латинских букв матрицы плана, определяемых как частные от деления соответствующих сумм квадратов на числа степеней свободы, с которой они оцениваются

SSA/(n-1); SSB/(n-1); SSC/(n-1); SSD/(n-1).

Проверить по критерию Фишера значимость оценок эффектов строк, столбцов и буквенных оценок

$$F_A = \frac{SS_A / (n-1)}{SS_{ост.} / \left[(n-1)\cdot(n-2)\right]}$$

Дисперсионный анализ результатов экспериментов (без повторных опытов)

Факторы дисперсионной модели	Число степеней свободы	Суммы квадратов эффектов	Средний квадрат	Расчетное значение F-критерия	Табличное значение F-критерия
Строки	$(n-1)$	SS_A	$MS_A = \frac{SS_A}{n-1}$	$F_A = \frac{MS_A}{MS_{ост}}$	
Столбцы	$(n-1)$	SS_B	$MS_B = \frac{SS_B}{n-1}$	$F_B = \frac{MS_B}{MS_{ост}}$	
Латинские буквы	$(n-1)$	SS_C	$MS_C = \frac{SS_C}{n-1}$	$F_C = \frac{MS_C}{MS_{ост}}$	
Остаток (суммарная ошибка)	$(n-1)\times(n-2)$	$SS_{ост}$	$MS_{ост} = \frac{SS_{ост}}{(n-1)(n-2)}$	—	
Итого . . .	(n^2-1)				

Дисперсионный анализ результатов экспериментов (с повторными опытами)

Факторы дисперсионной модели	Число степеней свободы	Суммы квадратов эффектов	Средний квадрат	Расчетное значение F-критерия	Табличное значение F-критерия
Строки	$(n-1)$	SS_A	$MS_A = \frac{SS_A}{n-1}$	$F_A = \frac{MS_A}{MS_{ош}}$	
Столбцы	$(n-1)$	SS_B	$MS_B = \frac{SS_B}{n-1}$	$F_B = \frac{MS_B}{MS_{ош}}$	
Латинские буквы	$(n-1)$	SS_C	$MS_C = \frac{SS_C}{n-1}$	$F_C = \frac{MS_C}{MS_{ош}}$	
Остаток	$(n-1) \times (n-2)$	$SS_{ост}$	$MS_{ост} = \frac{SS_{ост}}{(n-1)(n-2)}$	$F_{ост} = \frac{MS_{ост}}{MS_{ош}}$	
Ошибка (случайное рассеяние результатов опыта)	$n^2(m-1)$	$SS_{ош}$	$MS_{ош} = \frac{SS_{ош}}{n^2(m-1)}$	—	
Итого . . .	$n^2 m - 1$	$SS_{общ}$			

Пример

- Исследовали малоцикловую усталость (МЦУ) стали ВНЛ-3 в зависимости от чистоты обработки поверхности. $\overline{N}_F$
- Предварительные исследования показали, что, хотя с повышением чистоты поверхности средние значения долговечности образцов несколько увеличиваются, для возрастающих классов чистоты поверхности значения МЦУ , как правило, попадают в пределы доверительных интервалов оценки средних значений для более низких классов чистоты.
- Было отмечено также, что результаты испытания образцов из металла разных плавок, образцов, испытанных в разное время и на разных испытательных машинах, ложатся на кривые малоцикловой усталости с большим разбросом. Так, при σ_{max}= 800МПа разброс значений долговечности отдельных образцов составлял (суммарно по всем имевшимся результатам испытаний):
- Для МЦУ на базе тысячи циклов
- (19-59) циклов для точения, ∇4;
- (21-55) циклов для точения, ∇ 5;
- (40-60) циклов для точения, ∇6;
- (29-65) циклов для шлифования, ∇ 7.

Полученные результаты не могли быть объяснены только случайным рассеянием долговечности. В этой связи была поставлена задача произвести дисперсионный анализ результатов испытаний образцов, имеющих разную чистоту поверхности (X|). В качестве характерных источников неоднородности условий испытаний были выбраны:

уровень напряжений цикла X2;

плавка металла, X3;

испытательная машина , X4.

X1 – чистота поверхности образца (∇ 4 ∇5 ∇6 ∇7).

X2 – уровень напряжения цикла (нагрузка) 100, 90, 80, 70 кгс/мм2

X3 – плавка (A, B, C, D).

X4 – разные испытания машины одного класса (α, β, γ, δ).

Матрица плана эксперимента

X_2	X_1			
	Δ 4	Δ 5	Δ 6	Δ 7
100 90 80 70	Aγ Cδ Bβ Dα	Bδ Dγ Aα Cβ	Cα Aβ Dδ Bγ	Dβ Bα Cγ Aδ

Химический состав ВНЛ-3

Обозначение плавки в матрице плана	Содержание элементов, в %							
	C	Cr	Ni	Cu	Mo	Si	Mn	Nb
A	0,07	14,1	5,0	1,3	1,74	0,17	0,65	0,07
B	0,08	13,2	5,5	1,6	1,74	0,22	0,67	0,05
C	0,06	13,4	4,6	1,55	1,60	0,50	0,72	0,06
D	0,08	13,7	5,5	1,58	1,40	0,22	0,75	0,05

Средние значения результатов экспериментов, тыс. циклов

X_2	X_1			
	Δ 4	Δ 5	Δ 6	Δ 7
100	8,2	10,2	8,3	5,9
90	15,1	25,8	22,3	21,2
80	48,9	25,7	49,6	36,2
70	74,1	69,5	80,9	57,1

В каждой ячейке плана было испытано по 4 образца. Испытания проводились при рандомизации экспериментов.

Предварительные расчеты

$$SS_1 = \sum_{p=1}^{N} y_p^2 = (8,2 + 10,2 + 8,3 + ...)^2 = 29063,9$$

$$SS_2 = \frac{\sum_{i=1}^{N} A_i^2}{n} = \frac{1062,8 + 7123,4 + 25728,2 + 79298,6}{4} = 28303,2$$

$$SS_3 = \frac{\sum_{j=1}^{N} B_j^2}{n} = \frac{21403,7 + 17213,4 + 25953,2 + 14496,2}{4} = 19766,6$$

$$SS_5 = \frac{\sum_{l=1}^{N} D_l^2}{n} = \frac{16718,5 + 21491,6 + 22831,2 + 17424,0}{4} = 19616,3$$

$$SS_4 = \frac{\sum_{k=1}^{N} C_k^2}{n} = \frac{12836,9 + 25985,4 + 16666,8 + 24149,2}{4} = 19909,6$$

$$SS_{кор} = \frac{\left(\sum_{p=1}^{N} y_p\right)^2}{N} = \frac{559,0^2}{16} = 19530,1$$

Предварительные расчеты дисперсионного анализа

X_2	X_1				Итог	Квадрат А
	Δ 4	Δ 5	Δ 6	Δ 7		
100 90 80 70	8,2 15,1 48,9 74,1	10,2 25,8 25,7 69,5	8,3 22,3 49,6 80,9	5,9 21,2 36,2 57,1	32,6 84,4 160,4 281,6	$(32,6)^2$ $(84,4)^2$ $(160,4)^2$ $(281,6)^2$
Итого	146,3	131,2	161,1	120,4	559,0	–
Квадрат В	21403,7	17213,4	25953,2	14496,2		–
Эффект латинских букв						Квадрат С
A B C D	8,2 48,9 15,1 74,1	25,7 10,2 69,5 25,8	22,3 80,9 8,3 49,6	57,1 21,2 36,2 5,9	113,3 161,2 129,1 155,4	12836,9 25985,4 16666,8 24149,2
Эффект греческих букв.						Квадрат D
Α β γ δ	74,1 48,9 8,2 15,1	25,7 69,5 25,8 10,2	8,3 22,3 80,9 49,6	21,2 5,9 36,2 57,1	129,3 146,6 151,7 132	16718,5 21491,6 22831,2 17424,0

Вспомогательные расчеты

Суммы квадратов, характеризующие проверяемые эффекты, определены через вспомогательные суммы:

$$SS_{(строк)} = SS_2 - SS_{кор.} = 8773{,}2$$

$$SS_{(столбцов)} = SS_3 - SS_{кор.} = 236{,}6$$

$$SS_{(греч.букв)} = SS_5 - SS_{кор.} = 86{,}3$$

$$SS_{(лат.букв)} = SS_4 - SS_{кор.} = 379{,}5$$

$$SS_{(общ.)} = SS_1 - SS_{кор.} = 9533{,}9$$

$$SS_{ост.} = SS_{(общ.)} - \left[SS_{(строк)} + SS_{(столбцов)} + SS_{(лат.букв)} + SS_{(греч.букв)}\right] = 58{,}4$$

Дисперсионный анализ результатов

Исследуемый фактор	Число степеней свободы	Сумма квадратов	Средний квадрат*	$F_{критическое}$	
				$F_{расчетное}$	$F_{табличное}$ ($\alpha = 0,9$)
уровень нагрузки (строки)	3	8773	2924	150,3	5,39
чистота поверхности (столбцы)	3	236	78	4	5,39
плавка (латинские буквы)	3	379	126	6,5	5,39
испытательная машина (греческие буквы)	3	86	28	1,5	5,39
остаток	3	58	19		

Средний квадрат

$$M_i S_i = \frac{SS_i}{n-1}$$

Выводы

- Эффект от испытательных машин статистически не значим.
- Шероховатость поверхности может быть использована 4 или 5, т.к. эта сталь имеет невысокую чувствительность к чистоте поверхности.
- Уровень нагрузки статистически значим.
- Эффект от различия в плавках может перекрывать эффект от влияния чистоты поверхности.