

Дисперсионный анализ: задача с решением

ЗАДАНИЕ. В течение шести лет использовались четыре различных технологии по выращиванию сельскохозяйственной культуры. Необходимо установить влияние различных технологий на урожайность культуры по данным таблиц. Задачу просчитать вручную и на ПЭВМ.

Год	Технология(фактор А)			
	A1	A2	A3	A4
1	0,9	1	1,3	1,8
2	0,8	0,7	1,5	1,9
3	1	1,3	1,6	1,1
4	1,3	1	1,4	1,4
5	1,4	1,3	1,2	1,3
6	0,8	1,3	1,1	1,9

РЕШЕНИЕ.

Получим решение вручную.

Составляем расчетную таблицу для дисперсионного анализа.

i	A ₁		A ₂		A ₃		A ₄		Сумма
	y _{i1}	y _{i1} ²	y _{i2}	y _{i2} ²	y _{i3}	y _{i3} ²	y _{i4}	y _{i4} ²	
1	0,9	0,81	1	1	1,3	1,69	1,8	3,24	
2	0,8	0,64	0,7	0,49	1,5	2,25	1,9	3,61	
3	1	1	1,3	1,69	1,6	2,56	1,1	1,21	
4	1,3	1,69	1	1	1,4	1,96	1,4	1,96	
5	1,4	1,96	1,3	1,69	1,2	1,44	1,3	1,69	
6	0,8	0,64	1,3	1,69	1,1	1,21	1,9	3,61	
$T_j = \sum y_{ij}$	6,2		6,6		8,1		9,4		30,3
$S_j = \sum y_{ij}^2$		6,74		7,56		11,11		15,32	40,73
T_j^2	38,44		43,56		65,61		88,36		235,97

Найдем общую и факторную суммы квадратов отклонений, учитывая, что число уровней фактора $p = 4$, число испытаний на каждом уровне $q = 6$.

Получаем:

$$S_{\text{общ}} = \sum_{j=1}^p S_j - \frac{1}{pq} \left(\sum_{j=1}^p T_j \right)^2 = 40,73 - \frac{1}{24} 30,3^2 = 2,47625$$

$$S_{\text{факт}} = \frac{1}{q} \sum_{j=1}^p T_j^2 - \frac{1}{pq} \left(\sum_{j=1}^p T_j \right)^2 = \frac{1}{6} 235,97 - \frac{1}{24} 30,3^2 = 1,0746$$

Найдем остаточную сумму квадратов отклонений

$$S_{ост} = S_{общ} - S_{факт} = 2,47625 - 1,0746 \approx 1,402$$

Найдем дисперсии

$$s_{факт}^2 = \frac{S_{факт}}{p-1} = \frac{1,0746}{3} = 0,358$$

$$s_{ост}^2 = \frac{S_{ост}}{p(q-1)} = \frac{1,402}{4 \cdot 5} \approx 0,07$$

Сравним факторную и остаточную дисперсию с помощью критерия Фишера-Снедекора. Найдем наблюдаемое значение критерия

$$F_{набл} = \frac{s_{факт}^2}{s_{ост}^2} = \frac{0,358}{0,07} \approx 5,11.$$

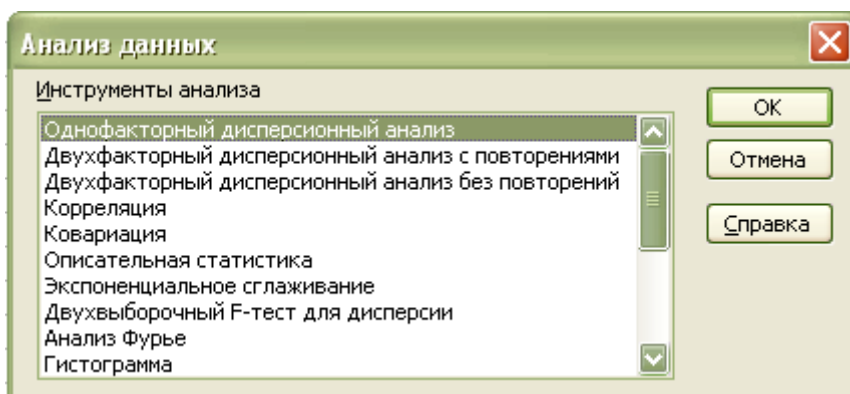
По числу степеней свободы $k_1 = p - 1 = 3$, $k_2 = p(q - 1) = 20$ и по уровню значимости $\alpha = 0,05 = 5\%$ находим критическую точку $F_{крит} = 3,098$. Так как $F_{набл} > F_{крит}$, следует отвергнуть нулевую гипотезу, влияние фактора (технологии) на урожайность значимо.

Решим задачу с помощью Excel.

Задаем таблицу данных.

Год	A1	A2	A3	A4
1	0,9	1	1,3	1,8
2	0,8	0,7	1,5	1,9
3	1	1,3	1,6	1,1
4	1,3	1	1,4	1,4
5	1,4	1,3	1,2	1,3
6	0,8	1,3	1,1	1,9

Запускаем надстройку Анализ данных, выбираем пункт Однофакторный дисперсионный анализ.



Задаем параметры анализа: данные и уровень значимости.

Однофакторный дисперсионный анализ

Входные данные

Входной интервал:

Группирование: по столбцам по строкам

Метки в первой строке

Альфа:

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

OK, Отмена, Справка

Получаем на выходе результаты:

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
Столбец 1	6	6,2	1,033333333	0,066666667		
Столбец 2	6	6,6	1,1	0,06		
Столбец 3	6	8,1	1,35	0,035		
Столбец 4	6	9,4	1,566666667	0,118666667		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	1,07458	3	0,35819444	5,110978993	0,008695578	3,098391224
Внутри групп	1,40167	20	0,07008333			
Итого	2,47625	23				

В таблице представлены средние значения и дисперсия для каждой из четырех выборок, наблюдаемая величина критерия F, критическое значение F и значимость F-статистики (вероятность). Также приведены факторная и остаточная сумма квадратов отклонений, числа степеней свободы для каждой из дисперсий и т.д.

Все данные совпадают с полученными в результате ручных расчетов в предыдущем пункте.

Так как значение вероятности $P = 0,00087$ меньше заданного уровня 0,05, можно сделать вывод, что влияние фактора значимо.