

Дисперсионный анализ

Докладчик: ординатор Хамаева А.А.

Дисперсионный анализ

```
graph TD; A[Дисперсионный анализ] --> B[Однофакторный]; A --> C[Двухфакторный]; A --> D[Многофакторный];
```

Однофакторный

Двухфакторный

Многофакторный

Основные понятия дисперсионного анализа

- Факторы – любые воздействия или состояния, определяющие ту или иную величину наблюдаемого признака
- Результативные признаки – наблюдаемые признаки, которые испытывают влияние изучаемых факторов
- Варианты – отдельные значения результативного признака

Фактор

ы

Организованные(контролируемые, основные)

Случайные(неконтролируемые, остальные)

Статистические комплексы (таблицы)

- Равномерные – с одинаковым числом значений в каждой клетке комбинационной таблицы
- Пропорциональные – число значений в различных клетках комбинационной таблицы различно, но соблюдена единая для всего комплекса пропорциональность между ними
- Непропорциональные – распределение значений по клеткам таблицы различно

-
- Общая дисперсия равна сумме дисперсий, вызванной организованными факторами (факториальной дисперсии) и дисперсии, вызванной случайными факторами (остаточной дисперсии)

$$C_o = C_{\phi} + C_c$$

C_o - общая дисперсия

C_{ϕ} - факториальная дисперсия

C_c - случайная дисперсия

- **Общая дисперсия:** $C_0 = \sum (X - X_c)^2$
- **Факториальная дисперсия:** $C_\phi = \sum (X_\phi - X_0)^2$
- **Случайная дисперсия:** $C_c = \sum (X - X_\phi)^2$

- Где X – отдельное значение результативного признака
- X_c – общая средняя арифметическая всего комплекса
- X_ϕ – групповая средняя

- Когда измеряется влияние нескольких факторов (в многофакторном комплексе), сумма дисперсий каждого из учитываемых факторов и случайной дисперсии должна быть равна общей дисперсии:

$$C_o = C_{\phi 1} + C_{\phi 2} + C_{\phi 3} + \dots + C_{\phi n} + C_c$$

- Доля участия отдельных факторов в формировании результаивного признака определяется из отношения групповых дисперсий к общей (в процентах):

$$\eta(\text{доля}) = C_{\phi} / C_o$$

- Для определения достоверности влияния факторов в группах с разным числом значений применяется так называемая *девиата*, т.е. дисперсия, приходящаяся на один элемент свободного варьирования или на одну степень свободы

$$D = C_o / n$$

- При вычислении общей девиаты:

$$D_o = C_o / n$$

- При вычислении факториальной девиаты: $D_\phi = C_\phi / n$

- При вычислении случайной девиаты:

$$D_c = C_c / n$$

Обработка однофакторного дисперсионного комплекса

Градации фактора	Обычная одежда	Синтетическая одежда	Комбинированная одежда
Варианты результативного признака	54 57	73 70	67 63

Вычисление общей дисперсии

- Сумма всех значений
 $54+57+73+70+67+63=384$
- Число значений $n=2 \times 3=6$
- Общая средняя $\bar{x} = 384/6=64$
- Общая дисперсия (сумма квадратов отклонений каждого значения от общей средней) $S = (54-64)^2 + (57-64)^2 + (73-64)^2 + (67-64)^2 + (63-64)^2 = 276$

Вычисление факториальной (групповой) дисперсии

- Определить средние величины результативного признака каждой градации фактора (X_{ϕ})
- Вычесть из них общую среднюю всего комплекса ($X - X_c$) и возвести в квадрат полученные отклонения
- Умножить их на повторность опыта $p^*(X_{\phi} - X_o)^2$ и сложить эти произведения

Вычисление случайной дисперсии

- Находятся групповые средние (X_{ϕ})
- Определяются отклонения каждой даты от своей групповой средней ($v - X_{\phi}$) и полученные отклонения возводятся в квадрат
- Квадраты отклонений складываются

- Отношение факториальной дисперсии к общей характеризует степень влияния изучаемого фактора:
 $S_f/S_o = 259 / 276 = 0,94$

Следовательно, статистическое влияние организованных факторов на результативный признак составляет 94%

- Степень влияния неучтенных факторов:
 $S_c/S_o = 17/276 = 0,06$

Доля влияния на результативный признак неучтенных факторов составляет 6%

Оценка достоверности влияния организованных и неучтенных факторов на величину результативного признака производится путем сравнения отношения факториальной и случайной девиат с соответствующими табличными значениями

Число степеней свободы в однофакторном комплексе определяется следующим образом:

- Для общей дисперсии $n_o = n - 1$, в примере $6 - 1 = 5$
- Для факториальной дисперсии $n_\phi = r - 1$, в примере $3 - 1 = 2$
- Для случайной дисперсии $n_c = n - r$, в примере $6 - 3 = 3$

Обработка двухфакторного дисперсионного комплекса

1. Вычисление общей дисперсии осуществляется как при однофакторном комплексе
2. Вычисление случайной дисперсии аналогично нахождению ее в однофакторном комплексе
3. Вычисление дисперсии суммарного действия организованных факторов

Двухфакторный статистический комплекс

Градации первого фактора	А1 пищевой рацион N 1		А 2 пищевой рацион N 2	
Градации второго фактора	В1 первая разновидность мышей	В2 вторая разновидность мышей	В1 первая разновидность мышей	В2 вторая разновидность мышей
Значение результивного признака	39 46 38	29 27 22	33 40 38	28 35 33

Вычисление факториальной дисперсии (фактор А)

Градации фактора А	А1	А2	N=12 $\Sigma_v = 408$
Значения по градациям фактора А	39 46 38 29 27 22	33 40 38 28 35 33	
Сумма значений по градациям фактора А	201	207	
Групповые средние по градациям фактора А (x_A)	33,5	34,5	$x_c = 34$
Отклонения групповых средних от общей средней (d_A)	-0,5	+0,5	
Квадраты отклонений (d_A^2)	0,25	0,25	$P_A = 6$ $C_A = 3$
Произведения квадратов отклонений на число значений градации фактора ($d_A^2 \times P_A = C_A \times P_A$)	1,5	1,5	

Вычисление факториальной дисперсии (фактор В)

Градация фактора В	Фактор В1	Фактор В2	
Значения по градациям фактора В	39 46 38 33 40 38	29 27 22 28 35 33	$N = 12$ $\Sigma_v = 408$ $X_c = 34$
Сумма значений по градациям фактора В (x_B)	234	174	
Групповые средние по градациям фактора В (x_B)	39	29	
Отклонения групповых средних от общей средней комплекса (d)	+5	-5	
Квадраты отклонений (d^2)	25	25	$P_B = 6$ $C_B = 300$
Дисперсия по фактору В ρ_B	150	150	

Вычисление факториальной дисперсии по сочетанию факторов

	Фактор A1		Фактор A2		
	Фактор B1	Фактор B2	Фактор B1	Фактор B2	
V	39 46 38	29 27 22	33 40 38	28 35 33	N=12
X _Ф	39 + 46 + 38 / 3 = 41	29 + 27 + 22 / 3 = 26	33 + 40 + 38 / 3 = 37	28 + 35 + 33 / 3 = 32	Σ _v = 408
X _A	39 + 46 + 38 + 29 + 27 + 22 / 6 = 33,5	39 + 46 + 38 + 29 + 27 + 22 / 6 = 33,5	33 + 40 + 38 + 28 + 35 + 33 / 6 = 34,5	33 + 40 + 38 + 28 + 35 + 33 / 6 = 34,5	X _c = 408 / 12 = 34
X _B	39 + 46 + 38 + 33 + 40 + 38 / 6 = 39	29 + 27 + 22 + 28 + 35 + 33 / 6 = 29	39 + 46 + 38 + 33 + 40 + 38 / 6 = 39	29 + 27 + 22 + 28 + 35 + 33 / 6 = 29	
C _O	+7	-8	+3	-2	P = 3
C _A	-0,5	-0,5	+0,5	+0,5	
C _B	+5	-5	+5	-5	
C _{AB}	+7 - (-0,5) - (+0,5) = +7 + 0,5 - 0,5 = 7	-8 - (+0,5) - (-5) = -8 - 0,5 + 5 = -3,5	+3 - (+0,5) - (+5) = +3 - 0,5 - 5 = -2,5	-2 - (+0,5) - (-5) = -2 - 0,5 + 5 = 2,5	C = 75
C _{AB} ^X P	6,25 × 3 = 18,75	6,25 × 3 = 18,75	6,25 × 3 = 18,75	6,25 × 3 = 18,75	

Определение достоверности значений

	Дисперсия (С)	Число степеней свободы (n)	Девиата (Д)	Отношение девиат (F _{вычисленное})	F(Табличное) при вероятностях	
					P = 0,95	P = 0,99
Фактор А (пища)	3	1	3,0	0,207	5,32	11,26
Фактор В (порода)	300	1	300,0	20,689	5,32	11,26
Сочетание АВ	75	1	75,0	5,172	5,32	11,26
Сумма А, В, АВ	378	3	126,0	8,689	4,11	7,60
Случайные факторы	116	8	14,5			
Итого	494	И	—	—	—	—

Спасибо за внимание!