ДИСПЕРСИОННЫЙ АНАЛИЗ

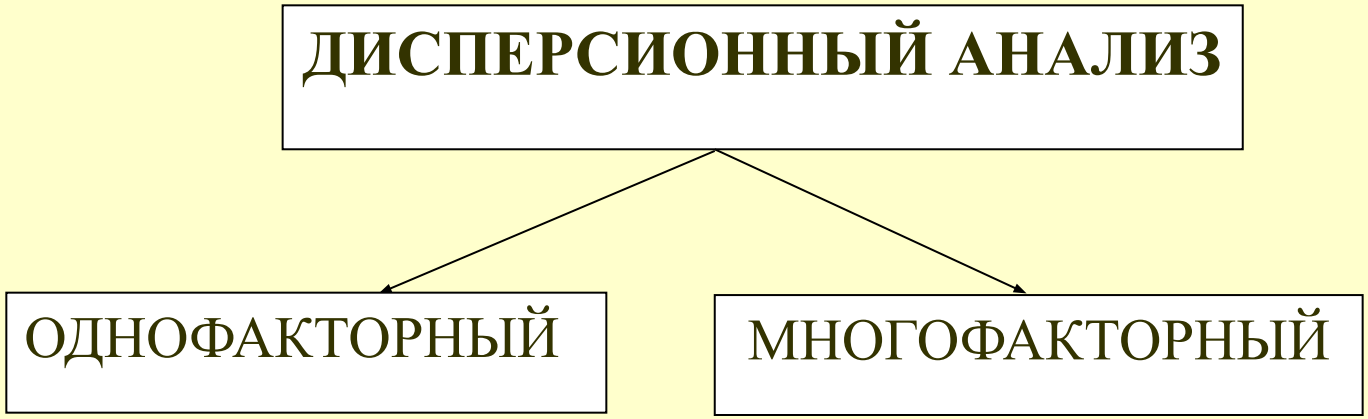
Лекция 7

План лекции:

- Виды дисперсионного анализа и его характеристики
- Этапы дисперсионного анализа
- Формулы для однофакторного дисперсионного анализа
- Сила влияния фактора
- Достоверность влияния фактора

Виды дисперсионного анализа и его характеристики

Раздел статистики, изучающий влияние факторов на изменчивость случайной величины, называется **дисперсионным анализом.**



Условия:

- изучаемые факторы должны быть независимыми;
- распределение выборочных данных должно соответствовать нормальному распределению или сводится к нему путем соответствующих преобразований

x – μ =A+e, где **μ** – средняя арифметическая генеральной совокупности;

x – конкретное значение переменной;

A – доля отклонения переменной, связанная с влиянием данного конкретного фактора;

e – остаточная часть отклонения, не объяснимая влиянием данного фактора.

- Признаки, изменяющиеся под влиянием тех или иных причин, называются результативными.
- Сами причины называются факторами.
- Конкретное числовое значение фактора называется градацией (или уровнем) фактора.
- Степень изменения всех признаков и отклонение их от средней арифметической ряда характеризуется дисперсией D(x):

$$D(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$D_{общ} = D_{факт} + D_{случ}$$

$$F = \frac{D_{факт}}{D_{случ}}$$ - критерий Фишера

Если **F>Fкр** (при вероятности P=0,95), то влияние фактора **существенно**. Если **F<Fкр** (вероятность P < 0,95), фактор **не влияет** на изучаемый признак.

Этапы дисперсионного анализа:

- Представить данные в виде таблицы.

Номер наблюдения (j)	**Уровни фактора(i)**					
	1	**2**	**3**	**. . .**	**I**	**a**
1	x_{11}	x_{21}	x_{31}		x_{i1}	x_{a1}
2	x_{12}	x_{22}	x_{32}		x_{i2}	x_{a2}
. . .						
j	x_{1j}	x_{2j}	x_{3j}		x_{ij}	x_{aj}
n	x_{1n}	x_{2n}	x_{3n}		x_{in}	x_{an}
Суммы по группам:	$\sum x_1$	$\sum x_2$	$\sum x_3$		$\sum x_i$	$\sum x_a$
Средние по группам:	$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$		$\overline{x}_i$	$\overline{x}_a$

i – индекс уровня фактора (от 1 до a);

j – индекс варианты (от 1 до n).

- Общее варьирование всех вариант (x_{ij}), независимо от того, в какой группе они находятся, вокруг общей средней $\overline{x}$ характеризуется дисперсией $\mathbf{D_{общ.}}$

$$D_{общ} = \frac{\sum_{ij} (x_{ij} - \overline{x})^2}{N - 1}$$

где **N=a·n**– число всех вариант;

$\mathbf{df_{общ.}}$ **= N–1** – число степеней свободы.

- Варьирование групповых средних $\overline{x}_i$ или средних каждого уровня данного изучаемого фактора вокруг общей средней $\overline{x}$, характеризуется факторной дисперсией $\mathbf{D_{факт}}$.

$$D_{факт} = \frac{\sum_i n_i(\overline{x}_i - \overline{x})^2}{a-1}$$

$\mathbf{df_{факт} = a - 1}$ – число степеней свободы.

n_i – среднее число вариант в каждой группе,

n – если число вариант в группах одинаково.

- Варьирование вариант x_{ij} внутри каждой группы вокруг каждой групповой средней $\overline{x}_i$ характеризует случайная или остаточная дисперсия $\mathbf{D_{случ}}$.

$$D_{случ} = \frac{\sum_i \left[\sum_j (x_{ij} - \overline{x}_i)^2 \right]}{N - a}$$

$df_{случ} = N - a$ – число степеней свободы.

Причем: **(N – a) + (a – 1) = N – 1**

Формулы для однофакторного дисперсионного анализа

Источник варьирования	**Сумма квадратов SS (числитель)**	**Число степеней свободы df (знаменатель)**	**Формулы для дисперсии MS**
Общее (все варианты)	$\sum_{ij}(x_{ij}-\bar{x})^2$	**N – 1**	$\frac{1}{N-1}\sum_{ij}(x_{ij}-\bar{x})^2$
Групповые средние (фактор А)	$\sum_{i} n_i(\bar{x}_i-\bar{x})^2$	**a –1**	$\frac{1}{a-1}\sum_{i} n_i(\bar{x}_i-\bar{x})^2$
Варианты внутри групп (случайные отклонения)	$\sum_{i}\left[\sum_{j}(x_{ij}-\bar{x}_i)^2\right]$	**N – a**	$\frac{1}{N-a}\sum_{i}\left[\sum_{j}(x_{ij}-\bar{x}_i)^2\right]$

Пример. Провести однофакторный дисперсионный анализ для выяснения влияния дозы удобрения (кг) на урожайность (ц/га).

№	A_1=15	A_2=20	A_3=25	A_4=30
	x_{1j}	x_{2j}	x_{3j}	x_{4j}
1	8	8,2	11,8	7,5
2	8,4	9	13	8,5
3	9	10	12,5	6,4
4	8,7	10	10	9
5	8,3	9,2	13,1	8,1
6	8,5	10	12	7
$\bar{x}_i$ =	8,5	9,1	12,1	7,8
$\sum x_i$ =	50,9	56,4	72,4	46,5
$\left(\sum x_i\right)^2$ =	2590,81	3180,96	5097,96	2162,25

$\sum x_{ij} = 226{,}2;\ \left(\sum x_{ij}\right)^2 = 226{,}2^2 = 51166{,}44\ \sum x_{ij}^2 = 2210{,}44$

Вычисления:

- Сумма квадратов $SS_{общ}$ для общей вариации:

$$SS_{общ} = \sum x_{ij}^2 - \frac{\left(\sum x_{ij}\right)^2}{N} = 2210{,}44 - \frac{51166{,}44}{24} = 78{,}505$$

- Сумма квадратов $SS_{факт}$ для вариации между группами:

$$SS_{фак} = \frac{1}{n}\sum\left(\sum x_i\right)^2 - \frac{\left(\sum x_{ij}\right)^2}{N} = \frac{1}{6}13175{,}78 - \frac{51166{,}44}{24} = 64.03$$

Средний квадрат, характеризующий факторную дисперсию $MS_{факт}$.:

$$MS_{фак} = \frac{SS_{фак}}{df_{фак}} = \frac{64{,}03}{3} = 31{,}34$$

- Сумма квадратов $SS_{случ}$ для вариации внутри групп:

$$SS_{сл}=SS_{общ} - SS_{фак} = 78,505 - 64,03=14,475$$

- Сумма квадратов $SS_{случ}$ для вариации внутри групп:

$$MS_{сл} = \frac{SS_{сл}}{df_{сл}} = \frac{14,475}{20} = 0,724$$

т.к $MS_{случ} < MS_{фак,}$ $$F = \frac{MS_{фак}}{MS_{сл}} = \frac{21,34}{0,724} = 29,475$$

а $F_{теор}=3,1$ для $P=0,95$ и $df_{сл}=20$ и $df_{фак}=3$

ВЛИЯНИЕ ФАКТОРА ДОСТОВЕРНО!

Сила влияния фактора

- Сила влияния фактора η_A^2 определяется:

$$\eta_A^2 = \frac{D_{факт.}}{D_{факт.} + D_{случ.}}$$ где

$$D_{факт} = \frac{MS_{факт.} - MS_{случ.}}{n} = \frac{MS_{факт.} - D_{случ.}}{n}$$

В нашем случае $$D_{факт} = \frac{21{,}34 - 0{,}724}{6} = 3{,}436$$

$$\eta_A^2 = \frac{D_{факт.}}{D_{факт.} + D_{случ.}} = \frac{3{,}436}{3{,}436 + 0{,}724} = 0{,}826$$

Вывод:

82,6% от действия всех факторов приходится на дозу удобрения, 17,4% – приходится на долю случайных факторов.

Для выявления наиболее эффективной дозы удобрения построим график

Достоверность влияния фактора

- Дисперсионный анализ позволяет установить, существуют ли **достоверные различия** между отдельными **уровнями** фактора.

$$S_d = \sqrt{\frac{D_{случ.}}{n}}$$

n – число вариант в каждой группе.

- Отношение разницы d к ее ошибке Sd , т.е. $t = \frac{d}{S_d}$, должно быть таким, чтобы оно гарантировало достоверность не менее чем при P=0,95.

- Коэффициент Q, рассчитан для разного количества **групп а** и степеней свободы **$df_{случ}$**.

$$S_d = \sqrt{\frac{D_{случ.}}{n}} = \sqrt{\frac{0,724}{6}} = 0,35;$$

d_{12}=9,1-8,5=0,6; d_{23}=12,1-9,1=3;

$$t_{12} = \frac{d_{12}}{S_d} = \frac{0,6}{0,35} = 1,7;$$

$$t_{23} = \frac{d_{23}}{S_d} = \frac{3}{0,35} = 8,57;$$

Q=4 для $df_{случ}$=20 и а=4;

t_{12}< Q, разница **не достоверна**!

t_{23}> Q, разница **достоверна**!

Вывод:

Внесение удобрений достоверно влияет на урожайность посевов. Наибольшую эффективность имеет фактор (удобрение), градация которого равна A_3=25 кг.

БЛАГОДАРЮ ЗА ВНИМАНИЕ