

6.1 Тема. Вычисление дисперсии однофакторного комплекса при малой выборке

Цель. Знакомство с методами вычисления дисперсии однофакторного комплекса

Дисперсионный комплекс называется однофакторным, если испытывается действие на признак одного регулируемого фактора. Однофакторные дисперсионные комплексы могут быть равномерными и неравномерными. Независимо от этого техника дисперсионного анализа однофакторных комплексов сводится главным образом к расчету показателей варьирования.

Для вычисления общей дисперсии (C_y) пользуются следующей формулой:

$$C_y = \sum v^2 - H, \quad H - \text{промежуточная величина, она равняется: } H = \frac{\sum (v^2)}{n}. \quad (53)$$

Остаточную дисперсию (C_z) вычисляют по формуле:

$$C_z = \sum v^2 - \sum h_x, \quad \text{где } \sum h_x = \frac{\sum (v^2)}{n}. \quad (54)$$

Факториальную дисперсию (C_x) вычисляют по формуле:

$$C_x = \sum h_x - H. \quad (55)$$

Пример. Влияние возраста матерей на живую массу телят при рождении. Порядок вычисления и необходимые данные приведены в таблице 6.1.1

Таблица 6.1.1 Обработка однофакторного комплекса при малой выборке

Показатели	Полновозрастные матери	Возраст матерей 31-36 мес.	Возраст матерей 25-30 мес	Σ
v (живая масса при рождении)	35, 36, 40, 38, 43, 42	38, 32, 40, 34, 35, 31	35, 37, 30, 31, 32	609
v^2	1225, 1296, 1600, 1444, 1849, 1764	1444, 1024, 1600, 1156, 1225, 961	1225, 1369, 900, 961, 1024	22067
N	6	6	5	17
$\sum v$	234	210	165	609
$(\sum v)^2$	$234^2=54756$	44100	27225	-
$\sum h_x = \frac{\sum (v^2)}{n}$	9126	7350	5445	21921

$\bar{X} = \frac{\sum v}{n}$	39	35	33	35,8
------------------------------	----	----	----	------

Для вычисления промежуточной величины H используются сводные показатели таблицы 6.1.1, строчки 4 и 3.

$$H = \frac{\sum (v^2)}{n} = \frac{609^2}{17} = 21817.$$

Дисперсии C_y , C_x , C_z вычисляются по вышеприведенным формулам, подставляя в них данные из таблицы:

$$C_y = 22067 - 21817 = 250;$$

$$C_x = 21912 - 21817 = 104;$$

$$C_z = 22067 - 21921 = 146.$$

Проверка правильности подсчетов производится суммированием: $C_y = C_x + C_z$, т.е. $104 + 146 = 250$. В данном случае подсчеты сделаны правильно.

Степень (доля) влияния разных факторов на варьирующий признак определяется отношением между дисперсиями C_x и C_y , C_z и C_y ; обозначают эти отношения через η^2 . Так, доля влияния учтенных факторов равняется $\eta^2_x = \frac{C_x}{C_z}$, а для неучтенных факторов $\eta^2_z = \frac{C_z}{C_y}$.

В нашем примере доля учтенных факторов равняется:

$$\eta^2_x = \frac{104}{250} = 0,415, \text{ или } 41,5 \%$$

Доля неучтенных факторов равняется:

$$\eta^2_z = \frac{146}{250} = 0,585, \text{ или } 58,5 \%$$

Достоверность факториальной дисперсии, то есть достоверно ли влияние и доля влияния фактора на изменчивость признака определяется коэффициентом достоверности Фишера (F). Для вычисления коэффициента Фишера необходимо определить число степеней свободы (v) и скорректированную дисперсию – девиату (σ^2).

Число степеней свободы для факториальной дисперсии (C_x) равно числу классов (l) по фактору минус единица.

$$v_x = l_x - 1; \text{ в нашем примере } = 3 - 1 = 2.$$

Для остаточной дисперсии (C_z) число степеней свободы равно численности выборки (n) минус число классов (l).

$$v_z = n - l_x; \text{ в нашем примере } = 17 - 3 = 14.$$

Число степеней свободы для общей дисперсии (C_y) равно численности выборки (n) без единицы.

$$v_z = n - 1; \text{ в нашем примере } = 17 - 1 = 16.$$

Корректированную дисперсию или девиату (σ^2) (факториальную и остаточную) вычисляют делением дисперсии на соответствующее число степеней свободы.

Факториальная девиата равна: $\sigma_x^2 = \frac{C_x}{v_x}$; в нашем примере $\sigma_x^2 = \frac{104}{2} = 52$.

Остаточная девиата равна: $\sigma_z^2 = \frac{C_z}{v_z}$; в нашем примере $\sigma_z^2 = \frac{146}{14} = 10,4$.

Коэффициент достоверности Фишера вычисляется делением факториальной скорректированной дисперсии (девиаты) на остаточную скорректированную дисперсию.

$F = \frac{\sigma_x^2}{\sigma_z^2}$; в нашем примере $F=5,5$.

Вычисленное значение F сравнивают с табличным значением F. Табличное значение F для данного примера при трех уровнях вероятности равно:

$F_{0,95}=3,7$; $F_{0,99}=6,5$; $F_{0,999}=11,8$.

В нашем примере вычисленное F равно 5,5, следовательно влияние возраста матерей на живую массу телят при рождении достоверно, при уровне вероятности $p=0,95$.

Задание 1. На четырех разновозрастных группах мужчин измерялась скорость кровотока в сосудах в 1 с. Результаты оказались следующие:

Возрастные группы мужчин	Варианты опыта (проб)			Средние (\bar{X}_i)
	1	2	3	
Первая	7	10	12	9,67
Вторая	9	7	14	10,00
Третья	11	16	20	15,67
четвертая	15	18	17	16,67

Определите, достоверны ли расхождения между средними показателями этих групп.

Задание 2. При изучении влияния светового режима на развитие гусениц дубового шелкопряда оказались следующие результаты наблюдений:

Варианты опыта	Количество гусениц к началу выкормки	Выживаемость гусениц на протяжении 5 дней					Средние групп
		1	2	3	4	5	
Контроль	150	9	8	7	8	17	9,8
Полная темнота	150	10	10	9	8	17	10,8

Свет 8 ч	150	9	8	8	9	16	10,0
Свет 12 ч	150	8	7	8	7	15	9,0
	150	9	8	8	7	17	9,4

Какой вывод следует сделать на основании этих данных?

Контрольные вопросы.

1. В чем заключается цель дисперсионного анализа?
2. Что называется общей, факториальной и остаточной дисперсией?
3. Какие бывают дисперсионные комплексы?