Неодинаковое число испытаний на различных уровнях

Выше число испытаний на различных уровнях предполагалось одинаковым. Пусть число испытаний на различных уровнях, вообще говоря, различно, а именно: произведено q_1 испытаний на уровне F_1, q_2 испытаний – на уровне $F_2, \dots, q_p$ испытаний – на уровне F_p. В этом случае общую сумму квадратов отклонений находят по формуле

$$S_{\text{общ}} = [P_1 + P_2 + \dots + P_p] - \left[\left(R_1 + R_2 + \dots + R_p\right)^3 / n\right],$$

где $P_1 = \sum_{i=l}^{q_1} x_{i2}^2$ – сумма квадратов наблюдавшихся значений признака

на уровне F_1;

$P_1 = \sum_{i=l}^{q_2} x_{i2}^2$ – сумма квадратов наблюдавшихся значений признака

на уровне F_1;

....

$P_p = \sum_{i=l}^{q_p} x_{i2}^3$ – сумма квадратов наблюдавшихся значений признака

на уровне F_1;

$R_1 = \sum_{i=l}^{q_1} x_{i1}$, $R_1 = \sum_{i=l}^{q_2} x_{i1}, \dots,$ $R_p = \sum_{i=l}^{q_p} x_{ip}$, – суммы наблюдавшихся

вычитали одно и то же число C и приняли $y_{ij} = x_{ij} - C$, то

значений признака соответственно на уровнях $F_1, F_2, \dots, F_p$;

$n = q_1 + q_2 + \dots + q_p$ – общее число испытаний (объем выборки.)

Если для упрощения вычислений из каждого наблюдавшегося значения x_{ij} вычитали одно и то же число *C* и приняли $y_{ij} = x_{ij} - C$, то

$$S_{\text{общ}} = [Q_1 + Q_2 + \dots + Q_p] - \left[\left(T_1 + T_2 + \dots + T_p\right)^2 / n\right],$$

$$Q_1 = \sum_{i=l}^{q_2} y_{i1}^2,\ Q_2 = \sum_{i=l}^{q2} y_{i2}^2, \dots, Q_p = \sum_{i=l}^{q_p} y_{ip}^2;\ T_1 = \sum_{i=l}^{q_q} y_{i1},\ T_2 = \sum_{i=l}^{q_2} y_{i2}, \dots, T_p = \sum_{i=l}^{q_p} y_{ip}.$$

где

Факторную сумму квадратов отклонений находят по формуле

$$S_{\text{факт}} = \left[(R_1^2/\, q_1) + (R_2^3/q_2) + \cdots + \left(R_p^3/q_p\right)\right] - \left[\left(R_1 + R_2 + \cdots + R_p\right)^2/n\right];$$

если значения признака были уменьшены $(y_{ij} = x_{ij} - C)$, то

$$S_{\text{факт}} = \left[(T_1^2/q_1) + (T_2^2/q_2) + \cdots \left(T_p^3/q_p\right)\right] - \left[\left(T_1 + T_2 + \cdots + T_p\right)^2/n\right].$$

Остальные вычисления производят, как и в случае одинакового числа испытаний:

$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}},$

$$S_{\text{факт}}^2 = \frac{S_{\text{факт}}}{(p-1)}, S_{\text{ост}}^2 = S_{\text{ост}}/(n-p).$$

Пример. Произведено 10 испытаний, из них 4 на первом уровне фактора, 4 – на втором и 2 – на третьем. Результаты испытаний приведены в таблице 5. Методом дисперсионного анализа при уровне значимости 0,01 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями.

Таблица 5

Номер испытания	Уровни фактора F_j		
i	F_1	F_2	F_3
1	40	62	92
2	44	80	76
3	48	71	
4	36	91	
$\bar{x}_{\text{гр}\, j}$	42	76	84

Решение. Для упрощения расчета вычтем $C = 64$ из каждого наблюдаемого значения: $y_{ij} = x_{ij} - 64$. Составим расчетную таблицу 6.

Используя таблицу 6, найдем общую и факторную суммы квадратов отклонений:

$$S_{\text{общ}} = \sum Q_1 - \left[\left(\sum T_j\right)^3 /n\right] = 3253 - [(-27)^2/10] = 3253 - 72{,}9 = \\ = 3180{,}1;$$

$$S_{\text{факт}} = [(T_1^2/q_1) + (T_2^2/q_2) + (T_3^2/q_3)] - \left[\left(\sum T_j\right)^2 /n\right] \\ = [(7744/4) + (441/4) + (1600/2)] - 72{,}90 = 2846{,}25 - 72{,}90 \\ = 2773{,}35.$$

Найдем остаточную сумму квадратов отклонений:

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}} = 3180{,}10 - 2773{,}35 = 406{,}75.$$

Найдем факторную и остаточную дисперсии:

$$S_{\text{факт}}^2 = \frac{S_{\text{факт}}}{(p-1)} = \frac{2773{,}35}{(3-1)} = \frac{2773{,}35}{2} = 1387;$$

$$S_{\text{ост}}^2 = \frac{S_{\text{ост}}}{(n-p)} = \frac{406{,}75}{(10-3)} = \frac{406{,}75}{7} = 58.$$

Сравним факторную и остаточную дисперсии по критерию F, для чего найдем наблюдаемое значение критерия:

$$F_{\text{набл}} = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} = \frac{1387}{58} = 23{,}9.$$

Т а б л и ц а 6

Номер испытания	Уровни фактора F_j						Итоговый столбец
i	F_1		F_2		F_3		
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	
1	-24	576	-2	4	28	784	
2	-20	400	16	256	12	144	
3	-16	256		49			
4	-28	784					

$Q_j = \sum y_{ij}^2$		2016		309		928	$\sum Q_j = 3253$
$T_j = \sum y_{ij}$	-88		21		40		$\sum T_j = -27$
T_j^2	7744		441		1600		

Учитывая, что число степеней свободы числителя $k_1 = 2,$ а знаменателя $k_2 = 7$ и уровень значимости $\alpha = 0{,}01,$ по таблице приложения 1 находим критическую точку: $F_{\text{кр}}\,(0{,}01; 2; 7) = 9{,}55.$

Так как $F_{\text{набл}} > F_{\text{кр}}$ – нулевую гипотезу о равенстве групповых средних отвергаем. Другими словами, групповые средние различаются значимо.