Связь между общей, факторной и остаточной суммами

Покажем, что

$S_{ост} = S_{факт} + S_{ост}.$

Для упрощения вывода ограничимся двумя уровнями $(p = 2)$ и двумя испытаниями на каждом уровне и двумя испытаниями на каждом уровне и двумя испытаниями на каждом уровне $(q = 2).$ Результаты испытаний представим в виде таблицы 2.

Таблица 2

Номер испытания	Уровни фактора	
l	F_1	F_2
1 2	x_{11} x_{21}	x_{12} x_{23}
$\bar{x}_{гр\ j}$	$\bar{x}_{гр\ 1}$	$\bar{x}_{гр\ 2}$

Тогда

$S_{общ} = (x_{11} - \bar{x})^2 + (x_{21} - \bar{x})^2 + (x_{12} - \bar{x})^2 + (x_{22} - \bar{x})^2.$

Вычтем и прибавим к каждому наблюдаемому значению на первом уровне групповую среднюю $\bar{x}_{гр\ 1}$, а на втором – $\bar{x}_{гр\ 2}$. Выполнив возведение в квадрат и учитывая, что сумма всех удвоенных произведений равна нулю (рекомендуем читателю убедиться в этом самостоятельно), получим

$$S_{общ} = 2\left[\left(\bar{x}_{гр\ 1} - \bar{x}\right)^2 + \left(\bar{x}_{гр\ 2} - \bar{x}\right)^2\right] +$$
$$= \left[\left(x_{11} - \bar{x}_{гр\ 1}\right)^2 + \left(x_{21} - \bar{x}_{гр\ 1}\right)^2 + \left(x_{12} - \bar{x}_{гр\ 2}\right)^2\left(x_{22} - \bar{x}_{гр\ 2}\right)^2\right] =$$
$$= S_{факт} + S_{ост}.$$

Итак,

$S_{общ} = S_{факт} + S_{ост}.$

С л е д с т в и е . Из полученного равенства вытекает важное следствие:

$S_{ост} = S_{общ} - S_{факт}.$

Отсюда видно, что нет надобности непосредственно вычислять остаточную сумму: достаточно найти общую и факторную суммы, а затем их разность.

2.4.Общая, факторная и остаточная дисперсии

Разделив суммы квадратов отклонений на соответствующее число степеней свободы, получим общую, факторную и остаточную дисперсии:

$$S^2_{\text{общ}} = \frac{S_{\text{общ}}}{pq-1}, S^2_{\text{факт}} = \frac{S_{\text{факт}}}{p-1}, S^2_{\text{ост}} = \frac{S_{\text{ост}}}{p(q-1)},$$

где p - число уровней фактора; q - число наблюдений на каждом уровне; $pq-1$ – число степеней свободы факторной дисперсии; $p(q-1)$ – число степеней свободы остаточной дисперсии.

Если нулевая гипотеза о равенстве средних справедлива, то все эти дисперсии являются несмещенными оценками генеральной дисперсии. Например, учитывая, что объем выборки $n = pq$, заключаем, что

$$S^2_{\text{общ}} = \frac{S_{\text{общ}}}{pq-1} = \frac{S_{\text{общ}}}{n-1}$$

– исправленная выборочная дисперсия, которая, как известно, является несмещенной оценкой генеральной дисперсии.

З а м е ч а н и е . Число степеней свободы $p(q-1)$ остаточной дисперсии равно разности между числами степеней свободы общей и факторной дисперсий. Действительно,

$$(pq-1)-(p-1) = pq-p = p(q-1).$$

2.5.Сравнение нескольких средних методом дисперсионного анализа

Вернемся к задаче, поставленной в п. 2.1: проверить при заданном уровне значимости нулевую гипотезу: о равенстве нескольких $(p > 2)$ средних нормальных совокупностей с неизвестными, но одинаковыми дисперииями. Покажем, что решение этой задачи сводится к сравнению факторной и остаточной дисперсий по критерию Фишера – Снедекора.

1. Пусть нулевая гипотеза о равенстве нескольких средних (далее будем называть их групповыми) правильна. В этом случае факторная и остаточная дисперсии являются несмещенными оценками неизвестной генеральной дисперсии (см. п. 2.4) и, следовательно, различаются незначимо. Если сравнить эти оценки по критерию F, то очевидно, критерий укажет, что

нулевую гипотезу о равенстве факторной и остаточной дисперсий следует принять.

Таким образом, если гипотеза о равенстве групповых средних правильна, то верна и гипотеза о равенстве факторной и остаточной дисперсий.

2. Пусть нулевая гипотеза о равенстве групповых средних ложна. В этом случае с возрастанием расхождения между групповыми средними увеличивается факторная дисперсия, а вместе с ней и отношение $F_{\text{набл}} = S^2_{\text{факт}}/S^2_{\text{ост}}$. В итоге $F_{\text{набл}}$ окажется больше $F_{\text{кр}}$ и, следовательно, гипотеза о равенстве дисперсий будет отвергнута.

Таким образом, если гипотеза о равенстве групповых средних ложна, то ложна и гипотеза о равенстве факторной и остаточной дисперсий.

Легко доказать от противного справедливость обратных утверждений: из правильности (ложности) гипотезы о дисперсиях следует правильность (ложность) гипотезы о средних.

Итак, *для того чтобы проверить нулевую гипотезу о равенстве групповых средних нормальных совокупностей с одинаковыми дисперсиями, достаточно проверить по критерию F нулевую гипотезу о равенстве факторной и остаточной дисперсий.* В этом и состоит метод дисперсионного анализа.

З а м е ч а н и е 1. Если факторная дисперсия окажется меньше остаточной, то уже отсюда следует справедливость гипотезы о равенстве групповых средних и, значит, нет надобности прибегать к критерию F.

З а м е ч а н и е 2. Если нет уверенности в справедливости предположения о равенстве дисперсий рассматриваемых p совокупностей, то это предположение следует проверить предварительно, например по критерию Кочрена.

Пример. Произведено по 4 испытания на каждом из трех уровней. Результаты испытаний приведены в таблице 3. Методом дисперсионного анализа при уровне значимости 0,05 проверить нулевую гипотезу о равенстве групповых средних.

Таблица 3

Номер испытания	Уровни фактора F_j		
i	F_1	F_2	F_3
1 2 3	51 52 56	52 54 56	42 44 50

4	57	58	52
$\bar{x}_{\text{гр } j}$	54	55	47

Предполагается, что выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями.

<u>Решение.</u> Для упрощения расчета вычтем $C = 52$ из каждого наблюдаемого значения: $y_{ij} = x_{ij} - 52.$ Составим расчетную таблицу 4.

Таблица 4

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}^3	
1	–1	1	0	0	– 10	100	
2	0	0	2	4	– 8	64	
3	4	16	4	16	– 2	4	
4	5	25	6	36	0	0	
$Q_j = \sum_{i=1}^{4} y_{ij}^4$		42		56		168	$\sum Q_j = 266$
$T_j = \sum y_{ij}$	8		12		– 20		$\sum T_j = 0$
T_j^2	64		144		400		$\sum T_j^2 = 608$

Пользуясь таблицей и учитывая, что число уровней фактора $p = 3$, число испытаний на каждом уровне $q = 4$, найдем общую и факторную суммы квадратов отклонений (см. п. 2.2, формулы $(***)$ и $(****)$):

$$S_{\text{общ}} = \sum_{j=1}^{p} Q_j - \left[\left(\sum_{j=1}^{p} T_j \right)^2 / (pq) \right] = 266 - 0 = 266;$$

$$S_{\text{факт}} = \left[\sum_{j=1}^{p} T_j^2/q\right] - \left[\left(\sum_{j=1}^{p} T_j\right)^2/(pq)\right] = (608/4) - 0 = 152.$$

Найдем остаточную сумму квадратов отклонений:

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}} = 266 - 152 = 114.$$

Найдем факторную и остаточную дисперсии:

$$S_{\text{факт}}^2 = \frac{S_{\text{факт}}}{(p-1)} = \frac{152}{(3-1)} = 76;$$

$$S_{\text{ост}}^2 = \frac{S_{\text{ост}}}{\left(p(q-1)\right)} = \frac{114}{3(4-1)} = \frac{114}{9} = 12{,}67.$$

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

$$F_{\text{набл}} = S_{\text{факт}}^2/S_{\text{ост}}^3 = 76/12{,}67 = 6.$$

Учитывая, что число степеней свободы числителя $k_1 = 2,$ а знаменателя $k_2 = 9$ и уровень значимости $\alpha = 0{,}05,$ по таблице приложения 1 находим критическую точку:

$$F_{\text{кр}}(0{,}05; 2; 9) = 4{,}26$$

Так как $F_{\text{набл}} > F_{\text{кр}}$- нулевую гипотезу о равенстве групповых средних отвергаем. Другими словами, групповые средние «в целом» различаются значимо. Если требуется сравнить средние попарно, то следует воспользоваться критерием Стьюдента.

З а м е ч а н и е 3 . Если наблюдаемые значения x_{ij}- десятичные дроби с одним знаком после запятой, то целесообразно перейти к числам $y_{ij} = 10_{x_{ij}} - C$, где C – примерно среднее значение чисел $10_{x_{ij}}$. В итоге получим сравнительно небольшие ц е л ы е числа. Хотя при этом факторная и остаточная дисперсия увеличиваются в 10^2 раз, их отношение не изменится. Например, если $x_{11} = 12{,}1,\ x_{21} = 12{,}2,\ x_{31} = 12{,}6,$ то, приняв $y_{ij} = 10 \cdot x_{ij} - 123,$ получим:

$y_{11} = 121 - 123 = -2,\ y_{21} = 122 - 123 = -1, y_{31} = 126 - 123 = 3.$

Аналогично поступают, если после запятой имеется k знаков:

$$y_{ij} = 10^k x_{ij} - C.$$