ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

УПРАВЛЕНИЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ И ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

Кафедра «Прикладная математика»

Учебно-методическое пособие

по дисциплине
«Математика»

«Однофакторный дисперсионный анализ»

Авторы
Рябых Г.Ю.,
Мул А.П.

Ростов-на-Дону, 2016

Аннотация

Однофакторный дисперсионный анализ.

Методические указания к практическим и лабораторным занятиям.

Методические указания включают краткое описание методов нахождения числовых Данные методические указания предназначены для самостоятельной работы студентов по ознакомлению и приобретению минимальных навыков практического использования одного из разделов математической статистики – дисперсионного анализа. Указания могут быть использованы на практических и лабораторных занятиях.

Предназначены для студентов всех направлений и специальностей.

Авторы

проф. Рябых Г.Ю.
ст. преподаватель Мул А.П.

Оглавление

ВВЕДЕНИЕ

Элементы дисперсионного анализа

Раздел математической статистики – дисперсионный анализ находит важные практические применения в различных областях научной, производственной и хозяйственной деятельности. Он является одним из вариантов факторного анализа данных наблюдений (опыта) при изучении влияния определенных факторов и их совокупностей на результаты наблюдений. Методами дисперсионного анализа можно решить задачи такого типа:

1. Установить, существенно ли влияние на величину У некоторого фактора Х, который может иметь несколько уровней X_1, X_2, …, X_p.

С помощью дисперсионного анализа может быть решен вопрос, подтверждают ли опытные данные или не подтверждают влияние Х на У, а также выявить наиболее высокий уровень Х.

2. Изучить воздействие на величину У несколько факторов.

$X^{(1)}$, $X^{(2)}$, …, $X^{(m)}$, каждый из которых может иметь несколько уровней:

$$X_1^{(1)}, X_2^{(1)}, \ldots, X_{K_1}^{(1)};$$

$$X_1^{(2)}, X_2^{(2)}, \ldots, X_{K_2}^{(2)};$$

$$X_1^{(m)}, X_2^{(m)}, \ldots, X_m^{(m)}.$$

Выявить наиболее существенно влияющие факторы, наилучшие их варианты, сравнить влияние отдельных факторов и уровней и их комбинаций.

Факторы могут быть как качественные (например, метод работы, тип станков, инструментов, вид материалов, вид удобрения и т.д.), так и количественные (например, давление, скорость резца, твердость металла, количество удобрения данного вида и т.д.).

Задача 1-ого типа является задачей однофакторного анализа.

<u>Пример 1.</u> Требуется проверить целесообразность введения нового технологического режима производства. Здесь У может, например, выражать средний процент брака за рабочую смену;

$X^{(1)}$ –вид технологического режима. Уровни $X^{(1)}$: $X_1^{(1)}$ – старый режим; $X_2^{(1)}$ – новый режим.

Вопрос о подтверждении преимущества нового технологического режима должен быть решен на основе анализа статического комплекса, содержащего значения (показатели) результативного признака (У), полученные путем проведенных испытаний. Он может быть представлен в виде комбинационной таблицы 1.

Таблица 1.

Номер испытания / Вид режима	1	2	3	….	P
$X_1^{(1)}$	$У_{11}$	$У_{12}$	$У_{13}$	….	$У_{1P}$
$X_2^{(1)}$	$У_{21}$	$У_{22}$	$У_{23}$	….	$У_{2P}$

P – число всех испытаний; $У_{ij}$ – значение У в j – том испытании при L – том уровне воздействующего фактора $X^{(1)}$ (i=1,2; j=1,2, …. , P).

Задача 2 – го типа относится к многофакторному дисперсионному анализу.

<u>Пример 2.</u> Пусть имеются практические основания предполагать, что средний процент брака У (см. пример № 1) зависит не только от варианта технологического режима производства (фактор $X^{(1)}$), но и от других факторов, например, от вида материала – фактор $X^{(2)}$ (предположим, что имеется 3 варианта материалов, которые могут быть использованы, т.е. $X^{(2)}$ имеет 3 уровня $X_1^{(2)}$, $X_2^{(2)}$, $X_3^{(2)}$), от двух различных типов станков – фактор $X^{(3)}$ ($X_1^{(3)}$, $X_2^{(3)}$ – уровни фактора $X^{(3)}$).

Проверка целесообразности вновь разработанного технологического режима может быть проведена путем анализа статического комплекса, представленного в виде комбинационной таблицы 2.

Таблица 2.

Вид технологич. режима	Тип станка	Виды материала		
		$X_1^{(3)}$	$X_2^{(3)}$	$X_3^{(3)}$
$X_1^{(1)}$	$X_1^{(2)}$	$У_{111}$	$У_{112}$	$У_{112}$
	$X_2^{(2)}$	$У_{121}$	$У_{122}$	$У_{123}$
$X_2^{(1)}$	$X_1^{(2)}$	$У_{211}$	$У_{212}$	$У_{213}$
	$X_2^{(2)}$	$У_{221}$	$У_{222}$	$У_{223}$

Здесь $У_{ijk}$ - процент брака в условиях i –го (i = 1, 2) технологического режима при использовании j –го (j = 1, 2) типа станков и К –го (К = 1, 2, 3) вида материалов.

Дисперсия σ^2 случайной величины (СВ) У является наиболее удобной и распространенной в статистике характеристикой меры вариации – меры изменчивости (колеблемости) СВ.

Основная идея дисперсионного анализа состоит в том, что общая сумма квадратов отклонений значений изучаемой величины от ее среднего значения, вычисляемая на основе всех данных наблюдения исследуемой величины (У), расщепляется на составные части, которые порождаются влиянием различных факторов и их комбинаций (факторные суммы) и не остаточную сумму, обусловленную не учитываемыми (случайными) факторами. Далее, с помощью этих сумм вычисляются соответственно общая, факторная и остаточная выборочная дисперсии. И исследованием значимости различия между этими дисперсиями решается вопрос о существенности влияния того или иного фактора или их комбинаций.

Мы остановимся здесь более подробно на однофакторном дисперсионном анализе.

ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Пусть СВ У может зависеть от фактора Х, имеющего q уровней X_1, X_2, $\ldots$, X_q. В результате эксперимента получено по р независимых наблюдений величины У на каждом из q уровней величины Х, которые выражаются величинами $У_{1j}$, $У_{2j}$, $\ldots$, $У_{pj}$; j = 1, 2, $\ldots$, q, т.е. всего pq значений величины У, представленных таблицей 3.

Таблица 3.

Номера наблюд.		Уровни X				
	X_1	X_2	$\ldots$	X_j	$\ldots$	X_q
1	$У_{11}$	$У_{12}$	$\ldots$	$У_{1j}$	$\ldots$	$У_{1q}$
2	$У_{21}$	$У_{22}$	$\ldots$	$У_{2j}$	$\ldots$	$У_{2q}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
i	$У_{i1}$	$У_{i2}$	$\ldots$	$У_{ij}$	$\ldots$	$У_{iq}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
p	$У_{p1}$	$У_{p2}$	$\ldots$	$У_{pj}$	$\ldots$	$У_{pq}$

Полагается, что независимые наблюдения $У_{ij}$ на каждом уровне взяты из нормальной генеральной совокупности о одной и той же дисперсией σ^2 на всех q уровнях.

Вычислим: среднее значение $\overline{y}$

$$\overline{y} = \frac{1}{pq} \sum_{i=1}^{p} \sum_{j=1}^{q} y_{ij} \qquad (1)$$

Общую сумму квадратов отклонений от среднего значения:

$$S_{общ} = \sum_{i=1}^{p} \sum_{j=1}^{q} (y_{ij} - \overline{y})^2 \qquad (2)$$

Факторную сумму квадратов:

$$S\text{ факт}= p\sum_{j=1}^{q}(\overline{\overline{Y}} - \overline{Y})^2 \qquad (3)$$

где $$\overline{Yj} = \frac{1}{p}\sum_{i=1}^{p} Yij \qquad (j=1,2,\ldots,q) \qquad (4)$$

среднее групповое значение Y , соответствующее группе наблюдений при постоянном j-том уровне фактора X.

$S_{общ}$ характеризует общую изменчивость – рассеяние (отклонение от среднего) величины Y, $S_{фак}$ характеризует изменчивость Y за счет фактора X (чем меньше влияние X, тем ближе групповые средние $\overline{Yj}$ и общей средней $\overline{Y}$, меньше S фак). Изменчивость, вызванная другими, не учитываемыми (случайными) факторами характеризуются величиной остаточной суммы квадратов отклонений

$$S\text{ ост}= \sum_{i=1}^{p}\sum_{j=1}^{q}(Yij - \overline{Y})^2 \qquad (5)$$

$S_{ост}$ отражает влияние случайных причин (отличных от фактора X).

Действительно, если бы на Y не оказывали влияния другие факторы, кроме X, то наблюдаемые при одном и том же j-ом уровне X значения Y теоретически не должны отличаться от $\overline{Yj}$ - среднего значения этой группы наблюдений; поскольку же на Y действуют и случайные факторы, отличные от X, то наблюдения при одном и том же уровне X различны – рассеяны вокруг групповой средней.

Покажем, что $S_{общ}= S_{фак}+S_{ост}$

(6)

$$S_{общ}= \sum_{i=1}^{p}\sum_{j=1}^{q}(Yij - \overline{Y})^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}(Yij - \overline{Yj})^2=$$

$$\sum_{i=1}^{p}\sum_{j=1}^{q}(Yij - \overline{Yj} + \overline{Yj} - \overline{Y})^2 =$$

$$\sum_{i=1}^{p}\sum_{j=1}^{q}(Yij - \overline{Yj})^2 + \sum_{i=1}^{p}\sum_{j=1}^{q}(\overline{Yj} - \overline{Y})^2 + 2\sum_{i=1}^{p}\sum_{j=1}^{q}(Yij - \overline{Yj})(\overline{Yj} - \overline{Y}) =$$

$S_{ост}+S_{фак}$,

так как $$\sum_{i=1}^{p}\sum_{j=1}^{q}(Yij-\overline{Y}j)(\overline{Y}j-\overline{Y})=0 \qquad (7)$$

Справедливость (7) можно проверить непосредственно:

$$\sum_{i=1}^{p}\sum_{j=1}^{q}(Yij-\overline{Y}j)(\overline{Y}j-\overline{Y})=\sum_{j=1}^{q}(\overline{Y}j-\overline{Y})\sum_{i=1}^{p}(Yij-\overline{Y}j)=\sum_{j=1}^{q}(\overline{Y}j-\overline{Y})(\sum_{i=1}^{p}Yij-\sum_{i=1}^{p}\overline{Y}j)=\sum_{j=1}^{q}(\overline{Y}j-\overline{Y})(\sum_{i=1}^{p}Yij-p\overline{Y}j)=$$

$$=\sum_{j=1}^{q}(\overline{Y}j-\overline{Y})(\sum_{i=1}^{p}Yij-p\frac{1}{p}\sum_{i=1}^{p}Yij)=0$$

Замечание 1. Для вычисления величин $S_{общ}$ и $S_{фак}$ удобно пользоваться формулами:

$$S_{общ}=\sum_{j=1}^{q}Pj-\frac{1}{pq}(\sum_{j=1}^{q}Rj)^{\wedge}2 \qquad (2^0)$$

$$S_{фак}=\frac{1}{p}\sum_{j=1}^{q}Rj^2-\frac{1}{pq}(\sum_{j=1}^{q}Rj)^{\wedge}2 \qquad (3^0)$$

где $Pj=\sum_{i=1}^{p}Yij^2$ - сумма квадратов значений величины Y на j-том уровне $Rj=\sum_{i=1}^{p}Yij$ - сумма значений величины Y на j-том уровне (j=1,2,……,q).

В дальнейшем под дисперсией будем понимать “исправленные” выборочные (вычисленные по данным наблюдений – выборке из генеральной совокупности), дисперсии, т.к. они являются несмещенными оценками генеральных

дисперсий (математическое ожидание “исправленной” выборочной дисперсии равно дисперсии генеральной).

Обозначим общую, факторную и остаточную дисперсии соответственно через σ^2 общ , σ^2 фак , σ^2 ост :

$$\sigma^2_{общ}=\frac{1}{n-1}*S_{общ}=\frac{1}{pq-1}*S_{общ} \qquad (8)$$

$$\sigma^2_{фак}=\frac{1}{q-1}*S_{фак} \qquad (9)$$

$$\sigma^2{}_{ост}=\frac{1}{q(p-1)}*S_{ост} \qquad (10),$$

Где $pq = n$ –общее число наблюдений ; q- число уровней фактора X;

р-число наблюдений на каждом уровне,

$pq - 1 = K_{общ}$; $q - 1 = K_{фак}$; q(p-1)= $K_{ост}$ - число степеней свободы соответственно общей, факторной и остаточной дисперсий

$K_{ост}= K_{общ} - K_{фак}$

Действительно, $K_{ост}= q(p-1)= pq-1-(q-1)= K_{общ} - K_{фак}$

Для того, чтобы решить вопрос о существенности влияния фактора X на изменение Y проверяют гипотезу: групповые средние равны.

Справедливость этой статистической гипотезы означает, что различие групповых средних незначимо и, следовательно M(Y) – математическое ожидание Y одно и то же на различных уровнях X, т.е. влияние последнего несущественно.

Заметим, что если гипотеза о равенстве групповых средних верна, то в этом случае каждая из дисперсий σ^2 общ , σ^2 фак , σ^2 ост является несмещенной оценкой неизвестной генеральной дисперсии. Поэтому факторная и остаточная дисперсии различаются незначимо. Если же эта гипотеза окажется несправедливой, то существенное различие в групповых средних влечет существенное различие факторной и остаточной дисперсий, т.е. можно считать, что опытные данные подтверждают существенность X на Y. Таким образом, решение нашей задачи сводится к проверке нулевой гипотезы, но различие между остаточной и факторной дисперсии незначимо.

Сравнение двух дисперсий нормальных генеральных совокупностей проводится с помощью F критерия (критерий Фишера), который является случайной величиной, равной отношению большей выборочной дисперсии к меньшей.

Так как при σ^2 фак < σ^2 ост можно считать расхождение между групповыми средними незначимым, то в этом случае нет надобности прибегать к критерию F ; (полагаем, что влияния фактора X несущественно. При σ^2 фак $\geq$ σ^2 ост вычисляем F

$$\text{табл} = \frac{\sigma^2 факт}{\sigma^2 ост}$$

Задаемся уровнем значимости λ. λ является вероятностью того, что в условиях справедливости нулевой гипотезы, критерий F примет значение, большее F кр , т.е. вероятность совершить ошибку 1-рода –отвергнуть правильную гипотезу. Уровень значимости λ берут обычно равным 0,05, или 0,01, или 0,001, что соответствует классификации явлений на редкие, очень редки и чрезвычайно редкие.

СВ F подчиняется определенному закону распределения вероятностей (закон Фишера-Снедекора). Этот закон позволяет при заданных степенях свободы $K1=K_{фак}$ и $K2=K_{ост}$

Вычислить вероятность того, что F(K1, K2) превзойдет заданное число F кр . По заданной вероятности, равной уровню значимости λ определяется F кр – критические точки распределения Фишера-Снедекора: F кр = F кр (λ ,K1,K2)

Имеются таблицы критических точек распределения F , из которых можно по λ , K1, K2 определить соответствующее F кр .

Если $F_{табл} < F_{кр}$, то нет оснований отвергнуть нулевую гипотезу, т.е. данные наблюдения не подтверждают существенность влияния фактора X на Y, если же $F_{табл} > F_{кр}$, то нулевую гипотезу отвергают, считается, что опытные данные подтверждают существенность влияния фактора X (расхождение между факторной и остаточной дисперсией значимо).

<u>Алгоритм однофакторного дисперсионного анализа:</u>

1. По данным наблюдений построить комбинационную таблицу (табл.3)
2. Вычислить общие и групповые средние (по формулам (1) и (4)).
3. Вычислить общую (S общ) и факторную (S фак) суммы квадратов (по формулам (2), (3)).
4. Вычислить $S_{ост} = S_{общ} - S_{фак}$

5. Если σ^2 фак $\geq$ σ^2 ост , то вычислить $F_{табл}$ = $\frac{\sigma^2 факт}{\sigma^2 ост}$

6. Определить F кр = F (λ ,K1,K2) по таблице критических точек распределения Фишера- Снедекора.

7. Сравнить $F_{табл}$ и $F_{пр}$.

Если $F_{табл}$ < $F_{пр}$ гипотеза Но принимается, расхождение между дисперсиями считается незначимым, в противном случае гипотеза отвергается, расхождение между дисперсиями значимо.

Замечание 2. Для упрощения расчетов иногда (при больших или дробных значениях Yij) удобно вместо Yij взять “уменьшенные ” значения Zij= Yij-C, где C равно общей средней $\overline{Y}$ или числу, близкому к $\overline{Y}$. Комбинационная таблица преобразуется в таблицу величин Zij . Расчет $S_{фак}$ и $S_{общ}$ следует вести по формулам (2^0) и (3^0), подставляя в них вместо Yij – Zij.

Замечание 3. Для выбора наилучшего уровня влияющего фактора (после того, как существенность его влияния подтверждена) можно сравнивать попарно групповые средние, выделить наилучшие варианты и далее сравнивать между собой эти варианты ; исследовать надежность различия средних в этих вариантах и т.д. При этом пользуются критерием Стьюдента.

Пример 4. Данные 4-х испытаний по признаку Y приведены в таблице 4. Исследовать существенность влияния величины X на Y.

Таблица 4.

Номер испытания	Уровни фактора X		
	X1	X2	X3
1	51	52	42
2	52	54	44

3	56	56	50
4	57	58	52

Полагая Zij= Yij -52 (c= 52) i= 1,2,3,4 ; j =1,2,3 составим расчетную таблицу 5.

Таблица 5.

Номер испытаний	Уровни фактора X						
	X1		X2		X3		
	Zi 1	(Zi1) ^2	Zi 2	(Zi2) ^2	Zi 3	(Zi3) ^2	
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	
Pj= $\sum_{i=1}^{4}(Xij)^\wedge 2$		42		56		168	$\sum_{j=1}^{3} Pj = 266$
RJ= $\sum_{i=1}^{4} Zij$	8		12		-20		$\sum_{j=1}^{3} Rj = 0$

(RJ)^2	64		144		400		$\sum_{j=1}^{3}(Rj)\hat{}2 = 608$

$$S_{общ} = \sum_{j=1}^{3} P_j - \frac{(\sum_{j=1}^{3} R_j)^2}{3*4} = 266 - 0 = 266$$

$$S_{фак} = \frac{1}{4}\sum_{j=1}^{3} R_j^2 - \frac{\sum_{j=1}^{3} R_j}{4*3} = \frac{608}{4} - 0 = 152$$

$$S_{ост} = 266 - 152 = 114$$

$$\delta_{фак}^2 = \frac{S_{фак}}{3-1} = \frac{152}{2} = 76$$

$$\delta_{ост}^2 = \frac{S_{ост}}{3*(4-1)} = \frac{114}{9} = 12{,}67\ ;$$

$$F_{набл} = \frac{\delta^2_{фак}}{\delta^2_{ост}} = \frac{76}{12{,}67} = 6.$$

По таблице критических точек распределения Фишера-Снедекора находим : при $K_1 = 3 - 1 = 2$; $K_2 = 3*(4 - 1) = 9$, $\alpha = 0{,}05$ находим $F_{табл} = F_{кр}(0{,}05; 2; 9) = 4{,}26$ (см., например, [2] стр. 420-421, табл. V).

Так как $F_{набл} > F_{кр}$, то нулевую гипотезу о равенстве дисперсии, а следовательно, гипотезу о равенстве групповых средних отвергаем. Таким образом, данные наблюдений подтверждают, что групповые средние различаются значимо, следовательно, подтверждается существенность влияния фактора X.

<u>Пример 5</u>. При изучении факторов производительности труда выделен один: X – образование рабочих. Установлены три его уровня: X_1 – начальное образование, X_2 – неполное среднее и X_3 – среднее. Из каждой группы отобрано по пять рабочих. Выработка рабочих (за один час в среднем) составила : с начальным образованием – 13, 14, 16, 18, 19; с неполным средним образованием – 19, 18, 21, 20, 22; со средним образованием 22, 23, 20, 21, 19.

Требуется исследовать существенность влияния образования Х рабочих на производительность (У) их труда.

Исходная таблица имеет вид:

Номер испыта-ния	Уровни фактора Х (образование)		
	X_1 – нач.	X_2 – н/сред.	X_3 – средн.
1	13	19	22
2	14	18	23
3	16	21	20
4	18	20	21
5	19	22	19

Положим с = 20 и составим расчетную таблицу:

<table>
<tr><td rowspan="3">Номер испытания</td><td colspan="6">Уровни фактора Х</td><td rowspan="8"></td></tr>
<tr><td colspan="2">X_1</td><td colspan="2">X_2</td><td colspan="2">X_3</td></tr>
<tr><td>X_{i1}</td><td>X_{i1}^2</td><td>X_{i2}</td><td>X_{i2}^2</td><td>X_{i3}</td><td>X_{i3}^2</td></tr>
<tr><td>1</td><td>-7</td><td>49</td><td>-1</td><td>1</td><td>2</td><td>4</td></tr>
<tr><td>2</td><td>-6</td><td>36</td><td>-2</td><td>4</td><td>3</td><td>9</td></tr>
<tr><td>3</td><td>-4</td><td>16</td><td>1</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>4</td><td>-2</td><td>4</td><td>0</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>5</td><td>-1</td><td>1</td><td>2</td><td>4</td><td>-1</td><td>1</td></tr>
<tr><td>$P_j = \sum_{i=1}^{5} x_{ij}^2$</td><td></td><td>106</td><td></td><td>10</td><td></td><td>15</td><td>$\sum_{j=1}^{3} P_j =$ 131</td></tr>
<tr><td>$R_j = \sum_{i=1}^{5} x_{ij}$</td><td>-20</td><td></td><td>0</td><td></td><td>5</td><td></td><td>$\sum_{j=1}^{3} R_j =$ 15</td></tr>
<tr><td>R_j^2</td><td>400</td><td></td><td>0</td><td></td><td>25</td><td></td><td>$\sum_{j=1}^{3} R_j^2$ = 425</td></tr>
</table>

$$S_{общ} = 131 - \frac{(-15)^2}{3*5} = 131 - 15 = 116\ ;$$

$$S_{фак} = \frac{1}{5}*425 - \frac{(-15)^2}{3*5} = 85 - 15 = 70\ ;$$

$S_{ост}$ = 116 – 70 = 46;

$$\delta_{\text{фак}}^2 = \frac{70}{3-1} = 35;$$

$$\delta_{\text{ост}}^2 = \frac{46}{3*(5-1)} = \frac{46}{12} = 3{,}84;$$

$$F_{\text{набл}} = \frac{35}{3{,}84} = 9{,}14.$$

По таблице критических точек распределения Фишера-Снедекора находим при K_1=3-1=2; K_2= 3*(5-1) = 12.

$F_{\text{табл}} = F_{\text{кр}}$ (0,05; 2; 12) = 3,88

Так как $F_{\text{набл}} > F_{\text{кр}}$, то гипотезу о равенстве групповых средних отвергаем. Данные наблюдений показывают существенность влияния фактора X (образование) на производительность труда рабочих.

Пример 6. Пробы из очень чистого железа, полученного двумя различными методами A и B, имели следующие точки плавления:

A	1493	1519	1518	1512	1512	1514
B	1509	1494	1512	1483	1507	1491

Существенно ли влияние на свойства металла метода его получения?

Положим с = 1500 и составим расчетную таблицу:

Номер испытания	Уровни фактора X – точки плавления				
	A		B		
	x_{i1}	x_{i1}^2	x_{i2}	x_{i2}^2	
1	-7	49	9	81	
2	19	361	-6	36	
3	18	324	12	144	
4	12	144	-17	289	
5	12	144	7	49	
6	14	196	-9	81	
$P_j = \sum_{i=1}^{6} x_y^2$		1218		680	$\sum_{j=1}^{2} P_j =$ 1898
$R_j = \sum_{i=1}^{6} x_{ij}$	68		-4		$\sum_{j=1}^{2} R_j =$ 64

R_j^2	4624		16		$\sum_{j=1}^{2} R_j^2 =$ 4640

$$S_{общ} = 1898 - \frac{64^2}{2*6} = 1898 - 341{,}3 = 1556{,}7$$

$$S_{фак} = \frac{1}{6} * 4640 - \frac{64^2}{12} = 773{,}3 - 341{,}3 = 432$$

$$S_{ост} = 1556{,}7 - 432 = 1124{,}7$$

$$\delta_{фак}^2 = \frac{432}{2-1} = 432$$

$$\delta_{ост}^2 = \frac{1124{,}7}{2*(6-1)} = 112{,}47$$

$$F_{набл} = \frac{432}{112{,}47} \approx 3{,}84$$

$F_{табл} = F_{кр}(0{,}05 ; 1 ; 10) = 4{,}96$

Так как $F_{набл} < F_{кр}$, то нулевую гипотезу принимаем. Влияние метода не существенно.

Решите самостоятельно следующие примеры:

Пример 7. Кислота непрерывным образом концентрируется на некотором типе оборудования, в результате чего часть оборудования ржавеет и со временем разрушается. Потери металла (в сотнях тонн) за период от установки оборудования до момента разрушения некоторой его части зафиксированы в таблице для трех линейных мастерских А, В и С. Проверить нулевую гипотезу,, по которой средняя продолжительность службы металла одна и та же для всех трех мастерских.

Мастерская	Потери металла				
A	84	60	40	47	34
B	67	92	95	40	98
C	46	93	100	92	92

Пример 8. Фруктовый сок хранился в течение нескольких месяцев в цистернах четырех типов, после чего комиссия из восьми человек определяла количество сока в каждой цистерне, причем, каждый дегустатор давал свою численную оценку по некоторой шкале. Результаты дегустации даны в таблице (см. ниже). А)

Проверить нулевую гипотезу, согласно которой тип цистерн, в которых хранится сок, не оказывает влияния на его качество; Б) Проверить нулевую гипотезу об отсутствии систематических расхождений между оценками дегустаторов.

Цистерна	Дегустатор							
	1	2	3	4	5	6	7	8
A	6,14	5,72	6,90	5,80	6,23	6,06	5,42	6,04
B	6,55	6,29	7,40	6,40	6,28	6,26	6,22	6,76
C	5,54	5,61	6,60	5,70	5,31	5,58	5,57	5,84
	4,81	5,09	6,61	5,03	5,15	5,05	5,77	6,17

Пример 9. По данным испытаний получена таблица значений результативного признака У при различных 5-ти уровнях некоторого фактора Х. Исследовать значимость влияния Х на У.

Номера испытаний	Уровни фактора Х				
	I	II	III	IV	V
1	83	81	76	78	79
2	61	61	67	67	64
3	78	71	75	72	74

Пример 10. Исследовать влияние фактора Х, имеющего 6 уровней, на некоторый фактор У, используя числовые характеристики У, полученные в результате 4-х испытаний и приведенные в таблице:

Номер испытания	Уровни Х					
	I	II	III	IV	V	VI
1	58	49	45	28	54	47
2	48	41	44	55	49	45
3	47	46	44	50	53	47
4	65	46	44	41	52	47

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Перечислите основные задачи, которые решаются с помощью дисперсионного анализа.
2. Какое различие между задачами, решаемыми с помощью однофакторного и с помощью многофакторного дисперсионного анализа?
3. В чем различие между общей и факторной суммами квадратов отклонения наблюдаемой величины от её среднего значения?
4. Как вычисляется остаточная сумма квадратов отклонений ($S_{ост}$) по $S_{фак}$ и $S_{общ}$?
5. Что характеризует $S_{ост}$, $S_{фак}$, $S_{общ}$?
6. В чем состоит основная идея дисперсионного анализа?
7. Опишите алгоритм однофакторного дисперсионного анализа.
8. Сформулируйте нулевую статическую гипотезу, проверяемую в процессе решения задач в однофакторном дисперсионном анализе? Какой статический критерий при этом используется?

ИСПОЛЬЗОВАНА ЛИТЕРАТУРА

[1.] ГМУРМАН В.Е. Теория вероятностей и математическая статистика. Изд-во “Высшая школа”, М. 1972 г., гл. XX, стр. 343-354.

[2.] ДЛИН А.М. Математическая статистика в технике. Изд-во “Сов. наука”, М., 1958г., гл. УII, стр. 201-222.