ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

УПРАВЛЕНИЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ И ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

Кафедра «Прикладная математика»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к проведению практических занятий
по дисциплине
«Математика»

«Двухфакторный дисперсионный анализ»

Авторы
Рябых Г.Ю.,
Мул А.П.

Ростов-на-Дону, 2016

Аннотация

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

Двухфакторный дисперсионный анализ

Методами двухфакторного дисперсионного анализа можно решить задачи такого типа:

1.Изучить воздействие на величину Y (функцию отклика) факторов $x^{(1)}$ и

$x^{(2)}$, каждый и которых имеет несколько уровней:

$x_1^{(1)}, x_2^{(1)}, \dots x_1^{(1)}; x_1^{(2)}, x_2^{(2)}, \dots x_k^{(2)}$.

2. Выявить наиболее существенно влияющие факторы, наилучшие их варианты, сравнивать влияние факторов и уровней и их комбинаций.

Пример

Пусть имеются практические основания предполагать, что средний процент брака Y зависит от выбора режима обработки $x^{(1)}$, ($x_1^{(1)}$ — старый режим, $x_2^{(1)}$ — новый режим) и от станка $x^{(2)}$ (станок $x_1^{(2)}$, $x_2^{(2)}$, $x_2^{(2)}$). Требуется проверить целесообразность

Вновь разработанного технологического режима, если проведено p испытаний.

Составим комбинационную таблицу:

Вид технологического режима	Тип станка	Номер испытания				
		1	2	3	...	p
$x_1^{(1)}$	$x_1^{(2)}$ $x_2^{(2)}$ $x_3^{(2)}$	y_{111} y_{121} y_{131}	y_{112} y_{122} y_{132}	y_{113} y_{123} y_{133}	... ...	y_{11p} y_{12p} y_{13p}

$x_2^{(1)}$	$x_1^{(2)}$ $x_2^{(2)}$ $x_3^{(2)}$	y_{211} y_{221} y_{231}	y_{212} y_{222} y_{232}	y_{213} y_{223} y_{233}	… … …	y_{21p} y_{22p} y_{23p}

Здесь y_{ihj} – процент брака в условиях i-го ($i=1,2$) технологического режима при использовании k-то (k=1,2,3) типа станков и при j-ом (j=1,2, …, p) испытании. В этом случае теоретико-вероятная схема имеет следующий вид: $y_{ikj} = a_{ik} + \varepsilon_{ikj}$, где a_{ik} — среднее значение результирующего признака при i-ом значении первого фактора и k-ом значении второго фактора, ε_{ijk} — независимые нормально распределенные величины. При этом,

$$a_{ik} = a + \alpha_i + \beta_k + \gamma_{ik},$$

a — генеральное среднее, α — главные эффекты первого фактора, β_k — главные эффекты второго фактора, γ_{ik} — эффекты взаимодействия.

Обозначим

$$\bar{y}_{ik} = \frac{1}{p}\sum_{j=1}^{p} y_{ikj}, \qquad \bar{y} = \frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K}\bar{y}_{ik}.$$

Общая вариация результирующего признака расчленяется на составляющие:

$$S_{общ} = \sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{j=1}^{p}(y_{ikj} - \bar{y})^2 = S_A + S_B + S_{AB} + S_{ост}.—$$

Здесь

$$S_A = Kp\sum_{i=1}^{I}\left(\frac{1}{K}\sum_{k=1}^{K}\bar{y}_{ik} - \bar{y}\right)^2;$$

$$S_B = Ip\sum_{k=1}^{K}\left(\frac{1}{I}\sum_{i=1}^{I}\bar{y}_{ik} - \bar{y}\right)^2;$$

$$S_{AB} = p\sum_{i=1}^{I}\sum_{k=1}^{K}\left(\bar{y}_{ik} - \frac{1}{K}\sum_{k=1}^{K}\bar{y}_{ik} - \frac{1}{I}\sum_{i=1}^{I}\bar{y}_{ik} + \bar{y}\right)^2;$$

$$S_{ост} = \sum_{i=1}^{I} \sum_{k=1}^{K} \sum_{j=1}^{p} (y_{ikj} - \overline{y}_{ik})^2 .$$

Выдвигаются гипотезы:

$H_A : \alpha_1 = 0$ — первый фактор влияет несущественно

$H_B : \beta_k = 0$ — второй фактор влияет несущественно

$H_{AB} : \gamma_{ik} = 0$ — первый и второй фактор влияет несущественно.

Для проверки гипотеы H_A о том, что первый фактор влияет несущественно, вычислим $S_A = K \cdot p \cdot \sum_{i=1}^{I} (\frac{1}{K} \sum_{k=1}^{K} (\overline{y}_{ik} + \overline{y})^2)$, здесь $\overline{y}_i = \frac{1}{K} \sum_{k=1}^{K} \overline{y}_{ik}$ — средняя по каждому уровню первого фактора. Для I уровня первого фактора $\overline{y}_1 = \frac{1}{3} \cdot (\overline{y}_{11} + \overline{y}_{12} + \overline{y}_{13}) = \frac{1}{3} (1.3 + 1.1 + 1.2) = 1.2$.

Для I I уровня первого фактора $\overline{y}_2 = 1.2$. Отсюда

$S_A = 3 \cdot 5 ((\overline{y}_2 - \overline{y})^2) = 15(0 + 0) = 0;$

$\sigma^2_{ф1} = \frac{S_A}{K_{ф1}} = 0$.

Для вычисления $S_{ост} = \sum_{i,k,j} (y_{ikj} - \overline{y}_{ik})^2$ составим комбинационную таблицу, содержащую $y_{ikj} - \overline{y}_{ik} = \Delta_{ikj}$ и $\Delta^2_{ikj} = \sum_{j=1}^{p} \Delta^2_{ikj}$.

		Δ_{ikj}					Δ^2_{ik}	
$x_1^{(1)}$	$x_1^{(2)}$	-0.1	0.1	0.2	-0.1	-0.1	0.08	$\bar{y}_{11}=1.3$
	$x_2^{(2)}$	0	-0.1	0.2	0	-0.1	0.06	$\bar{y}_{12}=1.1$
	$x_3^{(2)}$	0.1	-0.2	0	0.1	0	0.06	$\bar{y}_{13}=1.2$
$x_2^{(1)}$	$x_1^{(2)}$	-0.1	0.2	-0.1	-0.1	0.1	0.08	$\bar{y}_{21}=1.1$
	$x_2^{(2)}$	0.3	0.4	-0.3	0.2	0.2	0.42	$\bar{y}_{22}=1.2$
	$x_3^{(2)}$	0.1	0.1	0	0	-0.2	0.06	$\bar{y}_{23}=1.3$

$S_{ост} = \Delta^2_{11} + \Delta^2_{12} + \Delta^2_{13} + \Delta^2_{21} + \Delta^2_{22} + \Delta^2_{32} =$

$= 0.08 + 0.06 + 0.06 + 0.08 + 0.42 + 0.06 = 0.76$; $\sigma^2_{ост} = \frac{0.76}{24} \cong 0.03$.

Так как $\sigma^2_{ост} > \sigma^2_{ф1}$, то нет оснований отвергать гипотезу о том, что первый фактор влияет несущественно.

Обозначим $K_{ф1} = I - 1$, $K_{ф2} = K - 1$,

$K_{ф12} = (I-1)(K-1)$, $K_{ост} = I \cdot K \cdot (p-1)$, $\sigma^2_{ф1} = \frac{S_A}{K_{ф1}}$, $\sigma^2_{ф2} = \frac{S_B}{K_{ф2}}$, $\sigma^2_{ф12} = \frac{S_{AB}}{K_{ф12}}$, $\sigma^2_{ост} = \frac{S_{ост}}{K_{ост}}$.

Зададимся уровнем значимости α (α обычно берут равным 0,05; 0,01; 0,001).

A: а) если $\sigma^2_{ост} > \sigma^2_{ф1}$, то гипотеза H_A не отвергается

б) если $\sigma^2_{ф1} > \sigma^2_{ост}$, то применяется критерий Фишера:

$F_{набл} = \dfrac{\sigma^2_{ф1}}{\sigma^2_{ост}}$, $F_{кр} = F(\alpha, K_{ф1}, K_{ост})$ (значение $F_{кр}$ берутся из соответствующих таблиц). При $F_{набл} < F_{кр}$ нет оснований отвергать гипотезу H_A.

B: а) если $\sigma^2_{ост} > \sigma^2_{ф2}$, то гипотеза H_B не отвергается.

б) при $\sigma^2_{ф2} > \sigma^2_{ост}$ и $F_{набл} = \dfrac{\sigma^2_{ф2}}{\sigma^2_{ост}} < F_{кр} = F(\alpha, K_{ф2}, K_{ост})$ нет оснований отвергать H_B.

AB: а) если $\sigma^2_{ост} > \sigma^2_{ф12}$, то нет оснований отвергать гипотезу H_{AB}.

б) если $\sigma^2_{ф12} > \sigma^2_{ост}$ и $F_{набл} = \dfrac{\sigma^2_{ф12}}{\sigma^2_{ост}} < F_{кр} = F(\alpha, K_{ф12}, K_{ост})$, то гипотеза H_{AB} применяется.

Алгоритм двухфакторного дисперсионного анализа

1. По данным наблюдений построить комбинационную таблицу.
2. Вычислить групповые средние $\bar{y}_{ik}$ и общую среднюю $\bar{y}$.
3. Вычислить S_A, S_B, S_{AB}, $S_{ост}$.
4. Если $\sigma^2_{ост} > \sigma^2_{факт}$, то гипотеза принимается.

Если $\sigma^2_{ост} < \sigma^2_{факт}$, то сравниваем $F_{кр} = F(\alpha, K_{факт}, K_{ост})$ и $F_{набл} = \dfrac{\sigma^2_{факт}}{\sigma^2_{ост}}$. При $F_{набл} < F_{кр}$ гипотеза принимается.

Для примера по указанному алгоритму произведем расчет (I =2, K =3, p=5).

Режим	Тип станка	Номер испытания				
		1	2	3	4	5
$x_1^{(1)}$	$x_1^{(2)}$ $x_2^{(2)}$ $x_3^{(2)}$	1,2 1,1 1,3	1,4 1,0 1,0	1,5 1,3 1,2	1,2 1,1 1,3	1,2 1,0 1,2

$x_2^{(1)}$	$x_1^{(2)}$ $x_2^{(2)}$ $x_3^{(2}$	1,0 1,5 1,4	1,3 1,0 1,4	1,0 1,1 1,3	1,0 1,4 1,3	1,2 1,0 1,1

Найдем групповые средние (по каждой строке таблицы)

$$\bar{y}_{11} = \frac{1}{p}\sum_{j=1}^{p} y_{11j} = \frac{1}{5}(1.2 + 1.4 + 1.5 + 1.2 + 1.2) = 1.3;$$

$$\bar{y}_{12} = \frac{1}{p}\sum_{j=1}^{p} y_{12j} = \frac{1}{5}(1.1 + 1.0 + 1.3 + 1.1 + 1.0) = 1.1;$$

$\bar{y}_{13} = 1.2;\ \bar{y}_{21} = 1.1;\ \bar{y}_{22} = 1.2;\ \bar{y}_{23} = 1.3.$

Общее среднее

$$\bar{y} = \frac{1}{I \cdot K}\sum_{i=1}^{I}\sum_{k=1}^{K} \bar{y}_{ik} = \frac{1}{2 \cdot 3}(1.3 + 1.1 + 1.2 + 1.1 + 1.2 + 1.3) = 1.2.$$

Приведем несколько вариантов задач для самостоятельного решения. Во всех примерах проверить гипотезы: первый фактор влияет несущественно, влияние второго фактора несущественно, влияние совокупности первого и второго факторов несущественно(H_A,H_B, H_{AB})

№ 1. Пусть имеются практические основания предполагать, что средний процент брака зависит не только от технологического режима x_1, x_2, x_3 но и от вида материала $x_1^{(2)}$, $x_2^{(2)}$. Результаты опыта занесены в комбинационную таблицу (α=0.05).

		1	2	3	4	5
X 1 x_2 x_3 X_1	$X_1^{(2)}$ $X_2^{(2)}$	1.2 1.1	1.3 1.3	1.4 1.1	1.1 1.1	1.5 1.2
X_2	$X_1^{(2)}$ $X_2^{(2)}$	1.2 1.4	1.4 1.3	1.5 1.5	1.6 1.4	1.3 1.3

X_3	$X_1^{(2)}$ $X_2^{(2)}$	1.5 1.5	1.6 1.4	1.4 1.3	1.5 1.2	1.5 1.1

№2. При изучении фактров производительности труда выделены факторы $x^{(1)}$ образования рабочих и стаж работы $x^{(2)}$. Первый фактор имеет три уровня: начальное образование $x_1^{(1)}$, исполное среднее $x_2^{(1)}$ и среднее $x_3^{(1)}$. Второй фактор имеет два уровня: стаж работы более 5 лет ($x_1^{(2)}$) и стаж работы меньше 5 лет ($x_2^{(2)}$). Результаты обследования занесены в таблицу (α=0.10)

		1	2	3	4	5	6	7	8
$X_1^{(1)}$	$X_1^{(2)}$ $X_2^{(2)}$	1.3 1.1	1.2 1.2	1.3 1.1	1.4 1.0	1.2 1.0	1.1 1.1	1.3 1.1	1.2 1.2
$X_2^{(1)}$	$X_1^{(2)}$ $X_2^{(2)}$	1.3 1.2	1.3 1.3	1.4 1.2	1.3 1.3	1.4 1.4	1.2 1.2	1.1 1.0	1.2 1.0
$X_3^{(1)}$	$X_1^{(2)}$ $X_2^{(2)}$	1.5 1.4	1.6 1.4	1.7 1.3	1.5 1.3	1.4 1.2	1.4 1.1	1.3 1.2	1.1 1.3

№3. Пробы из чистого железа получили двумя методами А и В. Пробы были обработаны тремя способами. В комбинационной таблице даны точки плавления (α=0.05)

A	X_1 X_2 X_3	1493 1509 1512	1519 1494 1507	1518 1512 1514	1512 1483 1491
B	X_1 X_2 X_3	1499 1508 1514	1484 1512 1494	1491 1510 1507	1514 1514 1483

№4 Три литейные мастерские производят 2 вида изделий. Поте- ри металла в зависимости от мастерских A,B,C и вида изделий x_1 и x_2 даны в таблице (α=0.10)

A	X_1 X_2	80 98	60 92	40 40	47 30	33 35
B	X_1 X_2	67 46	93 92	35 92	45 40	40 50
C	X_1 X_2	84 46	60 93	40 92	34 30	42 34

№5 Сок, полученный из урожая двух виноградников A и B, хранился в четырех цистернах X_1, X_2, X_3, X_4. Дегустаторы поставили оценки сока, которые отражены в таблице (α=0.05).

A	X_1	6.1	5.7	6.9	5.8	6.2	6.1	5.4	6.0
	X_2	5.6	6.3	7.4	6.4	6.3	6.3	6.2	6.8
	X_3	5.5	5.6	5.7	5.3	5.6	5.6	5.8	6.6
B	X_3	5.5	5.6	5.7	5.3	5.6	5.6	5.8	6.6
	X_4	4.8	5.1	6.6	5.7	5.1	5.2	5.8	6.2

№6. Исследуется работа трех магазинов A, B, C, торгующих товарами двух видов: X_1 и X_2. Товарооборот за 6 месяцев представлен в таблице (α=0.10).

A	X_1	19	23	26	18	20	20
	X_2	18	35	20	20	32	27
B	X_1	40	24	22	18	16	15
	X_2	18	26	19	17	19	18
C	X_1	19	20	25	19	21	21
	X_2	18	35	18	20	32	24

№7. На заводе разработали два новых варианта технологического процесса. Испытывают старый вариант и 2 новых варианта: X_1, X_2, X_3. Испытания проводят на двух однотипных станках. Таблица колебаний производительности труда (α=0.10).

X_1	$X_1^{(2)}$	1.2	1.3	1.4	1.2	1.2	1.3
	$X_2^{(2)}$	1.3	1.4	1.1	1.0	1.4	1.0
X_2	$X_1^{(2)}$	1.4	1.5	1.6	1.7	1.6	1.6
	$X_2^{(2)}$	1.4	1.6	1.6	1.4	1.2	1.4
X_3	$X_1^{(2)}$	1.6	1.6	1.7	1.8	1.7	1.8
	$X_2^{(2)}$	1.6	1.5	1.7	1.8	1.7	1.9

№8. Партии мяса от пяти различных поставщиков загружаются в машину для упаковки в банки. Машина имеет четыре наполняющих цилиндра. Наудачу из каждой партии взято по 3 банки. Веса банок приведены в таблице (α=0.10). Здесь A, B,C, D, E – постановщики.

A	X_1	4	4	2
	X_2	4	3	5
	X_3	3	4	3
	X_4	4	2	3
B	X_1	6	1	2
	X_2	5	2	3
	X_3	4	4	4
	X_4	3	4	5
C	X_1	3	7	2
	X_2	1	8	4
	X_3	2	6	4
	X_4	3	3	5
D	X_1	3	0	6
	X_2	1	2	7
	X_3	2	4	5
	X_4	3	4	4
E	X_1	3	3	3
	X_2	0	4	8
	X_3	1	5	6
	X_4	3	2	7