Федеральное агентство по образованию

Государственное образовательное учреждение
высшего профессионального образования

Санкт-Петербургский государственный
технологический университет
растительных полимеров

ЭЛЕМЕНТЫ ДИСПЕРСИОННОГО АНАЛИЗА

Методические указания

для студентов всех специальностей дневной формы обучения

Санкт-Петербург
2009

УДК 519.23.237.4

Элементы дисперсионного анализа: методические указания для студентов всех специальностей дневной формы обучения / сост. З. Л. Абжандадзе, Е. В. Федорова, М. Э. Юдовин; ГОУВПО СПбГТУРП. - СПб., 2009. - 22 с.

В методических указаниях представлен необходимый теоретический материал. Даны рекомендации по решению задач. Предназначены для студентов всех специальностей дневной формы обучения.

Рецензент: доцент кафедры дифференциальных уравнений математико-механического факультета СПбГУ, канд. физ.-мат. наук В.В. Басов.

Подготовлены и рекомендованы к печати кафедрой высшей математики ГОУВПО Санкт-Петербургского государственного технологического университета растительных полимеров (протокол № 3 от 11 ноября 2008 г.).

Утверждены к изданию методической комиссией факультета промышленной энергетики ГОУВПО СПбГТУРП (протокол № 4 от 24 декабря 2008 г.).

Элементы дисперсионного анализа

Цель данных методических указаний – познакомить студентов с простейшими задачами, решаемыми средствами дисперсионного анализа, и помочь в выполнении индивидуального задания.

1. Основные задачи

Предположим, что изучается влияние одного или нескольких факторов на некоторую величину. Эти факторы могут принимать разные значения, называемые уровнями. Факторы могут быть как числовыми, так и нечисловыми. Например, на износ автомобильных покрышек может влиять как тип покрышки (нечисловой фактор), так и длина пробега (числовой фактор).

Вот некоторые из задач, которые ставятся в дисперсионном анализе:

- влияет ли некоторый фактор или группа факторов на изучаемую величину?
- какой из них имеет наибольшее влияние?
- зависит ли влияние факторов от их взаимодействия друг с другом?

2. Предварительные сведения

Напомним определения некоторых понятий из курса теории вероятностей и математической статистики, необходимых для понимания последующего материала:

а) Функция $F(x)$ называется <u>*функцией распределения*</u> случайной величины X, если для любого x выполняется равенство $F(x) = P(X < x)$, где $P(X < x)$ – вероятность попадания значения величины X в интервал $(-\infty; x)$.

б) Функция $f(x) = F'(x)$ называется <u>*плотностью распределения*</u>.

в) Числовые характеристики случайной величины:

$$\mu = M(X) = \int_{-\infty}^{+\infty} x f(x)dx$$ – <u>*математическое ожидание*</u>;

$$\sigma^2 = D(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$$ – <u>*дисперсия*</u>.

Математическое ожидание является в определенном смысле средним значением случайной величины, а дисперсия – характеристикой рассеяния значений случайной величины относительно ее среднего значения.

г) Число x_γ, определяемое уравнением $F(x_\gamma) = \gamma$, называется γ- <u>*квантилью*</u> распределения. Из определения следует, что γ- квантиль является возрастающей функцией от γ. Если график плотности симметричен относительно математического ожидания μ, то $F(\mu) = 0{,}5$ и, значит, в этом случае μ совпадает с $0{,}5$- квантилью.

д) <u>*Случайной выборкой*</u> объема n называется набор значений $(x_1,...,x_n)$ случайной величины, полученных в результате n независимых опытов. Эти значения называют в статистике наблюдениями.

е) Функция $K(x_1,...,x_n)$ от наблюдений называется <u>*несмещенной оценкой*</u> параметра a, если ее математическое ожидание равно a.

3. Однофакторный дисперсионный анализ

1. <u>Постановка задачи</u>

Пусть фактор A имеет m уровней и число y_{ij} получено в результате j-го опыта, проведенного на его i-м уровне, $j = 1,...,n_i; i = 1,...,m$. Числа y_{ij} называются наблюдениями, а n_i – число наблюдений, полученных на i-м уровне. Наблюдения представим в виде

$$y_{ij} = \mu_i + \varepsilon_{ij}, \qquad (1)$$

где μ_i - математическое ожидание y на i-м уровне, а ε_{ij} - случайная ошибка. Обычно наблюдения записывают в виде таблицы.

Таблица 1. Исходные данные

A_1	A_2	A_3	…	A_m
y_{11}	y_{21}	y_{31}	…	y_{m1}
y_{12}	y_{22}	y_{32}	…	y_{m2}
…	…	…	…	…
…	…	…	…	…
y_{1n_1}	y_{2n_2}	y_{3n_3}	…	y_{m,n_m}

Отметим, что столбцы в таблице могут быть разной длины, так как число наблюдений на разных уровнях фактора A не обязательно одинаково.

<u>*Пример 1*</u>. Четыре фирмы производят одинаковые изделия, y – некоторый показатель качества изделия (например, время безотказной работы). Здесь фактор A нечисловой – это фирма-производитель. Для сравнения качества изделий отбирают по 7 изделий у двух фирм и 9 и 8 изделий у двух других фирм и определяют значение y для каждого изделия. Получаем две случайные выборки объема 7 и две – объема 9 и 8. Здесь $m = 4$, $n_1 = 9$, $n_2 = n_3 = 7$, $n_4 = 8$. Требуется на основании этих данных выяснить, одинаково ли качество продукции у этих фирм, т.е. ответить на первый из перечисленных выше вопросов.

Если фактор не влияет на переменную y, то рассеяние ее значений вызвано лишь случайными ошибками, а математические ожидания на всех уровнях одинаковы. В терминах математической статистики задача сводится к проверке гипотезы $H_o : \mu_1 = \mu_2 = ... = \mu_m$.

Обозначим $\alpha_i = \mu_i - \mu$, $\mu = \frac{1}{n}\sum_{i=1}^{m} n_i \mu_i$. Число α_i называется эффектом фактора A на i-м уровне. Тогда уравнение (1) и гипотеза H_o принимают вид

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{2}$$

$$H_o : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0\ . \tag{3}$$

Далее предполагается, что случайные ошибки удовлетворяют следующим условиям:

а) имеют нулевое математическое ожидание;

б) имеют постоянную дисперсию, т.е. не зависящую ни от уровня фактора, ни от номера наблюдения;

в) подчиняются нормальному распределению.

2. <u>Оценки параметров модели (2)</u>

Определим следующие величины:

$y_{ig} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, $i = 1,2,\ldots,m$ – средние значения по столбцам;

$e_{ij} = y_{ij} - y_{ig}$ – отклонения от среднего в каждом столбце;

$y_{gg} = \frac{1}{n}\sum_{i,j} y_{ij}$ – общее среднее, $n = \sum_{i=1}^{m} n_i$;

$\hat{\alpha}_i = y_{ig} - y_{gg}$ – отклонения средних по столбцам от общего среднего;

Если выполнены допущения а), б), в), то можно доказать, что

$$M(e_{ij}) = 0,\ M(y_{gg}) = \mu,\ M(\hat{\alpha}_i) = \mu_i - \mu = \alpha_i, \tag{4}$$

где $\mu = \frac{1}{n}\sum_{i=1}^{m} n_i \mu_i$.

На языке математической статистики соотношения (4) означают, что случайные величины y_{ig} и $\hat{\alpha}_i$ являются несмещенными оценками параметров μ_i и α_i.

3. <u>Идея проверки гипотезы (3)</u>

Вычислим следующие суммы квадратов:

$S_П = \sum_{i,j}(y_{ij} - y_{gg})^2$ – полная сумма квадратов;

$S_M = \sum_{i=1}^{m} n_i (y_{ig} - y_{gg})^2 = \sum_{i=1}^{m} n_i \hat{\alpha}_i^2$ – межгрупповая сумма квадратов;

$S_B = \sum_{i,j}(y_{ij} - y_{ig})^2$ – внутригрупповая сумма квадратов.

Справедливо соотношение

$$S_П = S_M + S_B. \tag{5}$$

Здесь S_M характеризует рассеяние средних по столбцам относительно общего среднего, т.е. рассеяние между группами (уровнями фактора), а S_B характеризует рассеяние значений y_{ij} относительно y_{ig}, т.е. рассеяние внутри групп (столбцов таблицы).

Метод проверки гипотезы (3) основан на следующей идее. Если гипотеза верна, т.е. $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$, то величины $\hat{\alpha}_i$ должны быть достаточно близки к 0. Тогда вклад S_M в $S_П$ по сравнению с S_B должен быть мал. Поэтому малое значение S_M является доводом в пользу гипотезы, а большое значение S_M является доводом против гипотезы. Очевидно, в этом рассуждении не хватает точного указания, какое значение S_M считать малым.

4. <u>Применение F - критерия для проверки гипотезы</u>

Опишем точный метод проверки гипотезы (3), основанный на F - критерии.

1. Вычисляем средние суммы квадратов:

$$s_M^2 = S_M / (m-1);\quad s_B^2 = S_B / (n-m).$$

Числа (m – 1) и (n – m), на которые делятся суммы квадратов, называются *<u>степенями свободы.</u>*

2. Вычисляем значение F - критерия

$$F_{расч} = \frac{s_M^2}{s_B^2}.$$

3. Задаем число $0 < \alpha < 1$ и из таблицы квантилей F - распределения со степенями свободы $(m-1), (n-m)$ при уровне значимости α находим критическое значение $F_{крит}$.

Правило:

если $F_{расч} > F_{крит}$, то гипотеза отвергается;

если $F_{расч} \le F_{крит}$, то гипотеза принимается.

Замечания.

1) Вероятностный смысл α состоит в следующем. Предположим, что гипотеза H_0 верна, но из-за случайных ошибок вычисленное значение F оказалось больше критического, т.е. $F_{расч} > F_{крит}$. Тогда согласно сформулированному выше правилу мы должны отвергнуть H_0, хотя на самом деле она верна. Получается, что, применяя это правило, мы в этом случае совершим ошибку, называемую ошибкой 1-го рода (отвергается верная гипотеза). Вероятность такой ошибки равна вероятности неравенства $F_{расч} > F_{крит}$, вычисленной в предположении верности гипотезы H_0, т.е. равна α.

2) $F_{крит}$ зависит от выбранного значения α, причем $F_{крит}$ увеличивается при уменьшении α. Поэтому, уменьшая α, всегда можно добиться выполнения неравенства $F_{расч} \le F_{крит}$ и тем самым принятия гипотезы. Однако, уменьшая α, мы увеличиваем β – вероятность ошибки 2-го рода: принять H_0, когда на самом деле она неверна. Обычно используют $\alpha = 0,05; 0,025; 0,01$. Задать значение β мы не можем, так как оно зависит от неизвестных нам истинных значений эффектов α_i.

Пример 2.

Таблица 2. Исходные данные к примеру 2

Номер наблюдения	A_1	A_2	A_3	A_4
1	9,57	11,17	12,07	13,12
2	8,33	10,81	11,06	10,81
3	10,13	11,73	10,90	12,36
4	10,29	10,41	10,17	12,75
5	8,85	13,18	11,29	9,91
6	11,19	10,86	9,66	10,06
7	11,19	11,11	11,71	12,07
8	9,96	-	-	11,10
9	10,33	-	-	-
y_{ig}	9,98	11,32	10,98	11,52

Здесь $m = 4, n_1 = 9, n_2 = 7, n_3 = 7, n_4 = 8$.

Из таблицы видно, что средние по столбцам заметно различаются. Однако нельзя исключить, что это различие вызвано лишь случайным рассеянием данных, в то время как "истинные" значения средних, т.е. μ_i, одинаковы. Для проверки гипотезы H_0 применим описанный выше метод. Результаты расчетов приведены в таблице 3.

Таблица 3. Результат дисперсионного анализа

Источник рассеяния	Сумма квадратов	Степени свободы	Средняя сумма квадратов	$F_{расч} = \frac{s_M^2}{s_B^2}$	p
между группами	12,003	3	4,001	3,99	0,018

Окончание табл. 3

Источник рассеяния	Сумма квадратов	Степени свободы	Средняя сумма квадратов	$F_{расч} = \frac{s_M^2}{s_B^2}$	p
внутри групп	27,047	27	1,002	-	-
полная	39,05	30	-	-	-

Поясним содержание таблицы. Второй столбец содержит суммы квадратов $S_M; S_B; S_П$, смысл которых указан в первом столбце; в 3-м столбце – степени свободы, равные $(m - 1)$, $(n - m)$ и $(n - 1)$ соответственно; 4-й столбец получается делением сумм квадратов на их степени свободы. В последний столбец обычно помещают вероятность $p = P(F > F_{расч})$. Дело в том, что для проверки неравенства

$$F_{расч} > F_{крит} \tag{6}$$

потребуется сначала найти $F_{крит}$, а для этого нужна таблица квантилей F-распределения, которая не всегда доступна. Заметим, что $\alpha = 1 - \Phi(F_{крит})$, $p = 1 - \Phi(F_{расч})$, где $\Phi(F)$ – функция распределения Фишера. Функция $\Phi(F)$ возрастающая, поэтому неравенство (6) равносильно (7)

$$p < \alpha. \tag{7}$$

Поэтому вместо неравенства (6) можно пользоваться неравенством (7). В данном примере при $\alpha = 0,05$ получаем $p = 0,018 < 0,05 \Rightarrow H_o$ принимается на уровне значимости 0,05.

4. Двухфакторный дисперсионный анализ с однократными наблюдениями

Здесь рассматривается возможное влияние нескольких факторов на некоторую переменную y. Как и в однофакторном анализе, факторы могут быть и числовыми, но и в этом случае учитывается только, сколько различных значений (уровней) принимает фактор, а не величина этих значений. Рассмотрим наиболее простую двухфакторную модель.

Изучаем влияние факторов A и B на переменную y. Модель можно представить в виде

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}, \quad i = 1,\dots,I; j = 1,\dots,J, \tag{8}$$

где μ_{ij} – некоторые константы, ε_{ij} – случайные ошибки, I, J – число уровней факторов A и B соответственно, а y_{ij} – наблюдение, полученное на i-м уровне фактора A и j-м уровне фактора B.

Случайные ошибки удовлетворяют тем же требованиям, что и в предыдущем пункте.

Предположим, что между факторами нет взаимодействия. Это означает, что влияние одного фактора на величину y не зависит от того, на каком уровне находится другой фактор. От этого допущения нельзя отказаться, так как при однократных наблюдениях слишком мало исходных данных для оценки более сложной модели. Тогда уравнение (8) можно записать в виде

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1,\dots,I; j = 1,\dots,J. \tag{9}$$

Здесь μ, α_i, β_j – некоторые константы, имеющие следующий смысл: μ – общее среднее, α_i – эффект фактора A на его i-м уровне, β_j – эффект фактора B на его j-м уровне.

Пусть мы хотим проверить предположение о том, что фактор A или B не влияет на переменную y. Иначе говоря, нужно проверить гипотезы о равенстве нулю соответствующих эффектов

$$H_A : \alpha_i = 0, i = 1,\dots,I \quad \text{и} \quad H_B : \beta_j = 0, j = 1,\dots,J.$$

Для этого нужно вычислить следующие величины:

$y_{i\text{g}} = \frac{1}{J}\sum_{j=1}^{J} y_{ij}$ – средние по строкам;

$y_{\text{g}j} = \frac{1}{I}\sum_{i=1}^{I} y_{ij}$ – средние по столбцам;

$y_{\text{gg}} = \frac{1}{IJ}\sum_{i,j} y_{ij}$ – общее среднее.

Эти величины вместе с исходными данными обычно записывают в таблицу 4.

Таблица 4. Исходные данные

j / i	1	2	…	J	Средние по строке
1	y_{11}	y_{12}	…	y_{1J}	$y_{1\text{g}}$
2	y_{21}	y_{22}	…	y_{2J}	$y_{1\text{g}}$
…	…	…	…	…	
I	y_{I1}	y_{I2}	…	y_{IJ}	$y_{I\text{g}}$
Средние по столбцу	$y_{\text{g}1}$	$y_{\text{g}2}$	…	$y_{\text{g}J}$	y_{gg}

Далее вычисляем $\hat{\alpha}_i = y_{i\text{g}} - y_{\text{gg}}$; $\hat{\beta}_j = y_{\text{g}j} - y_{\text{gg}}$. Эти величины являются несмещенными оценками соответствующих параметров α_i, β_j модели (9). Поэтому, если гипотезы H_A, H_B верны (т.е. эффекты равны нулю), то $\hat{\alpha}_i, \hat{\beta}_j$ не должны значимо отличаться от нуля.

Алгоритм проверки гипотез, как и в предыдущем пункте, основан на F - критерии и состоит в следующем. Вычисляем

а) суммы квадратов

$$S_A = J\sum_{i=1}^{m_A}\hat{\alpha}_i^2; \quad S_B = I\sum_{j=1}^{m_B}\hat{\beta}_j^2; \quad S_П = \sum_{i,j}(y_{ij} - y_{\text{gg}})^2; \quad S_R = S_П - S_A - S_B.$$

Здесь S_A, S_B характеризуют величину эффектов, $S_П$ – полная сумма квадратов, а остаточная сумма S_R служит для оценки дисперсии ошибок наблюдений при условиях а), б), в) из предыдущего пункта;

б) средние суммы квадратов:

$$s_A^2 = \frac{S_A}{I-1}; \quad s_B^2 = \frac{S_B}{J-1}; \quad s_R^2 = \frac{S_R}{(I-1)(J-1)};$$

в) значения F - критериев для гипотез H_A, H_B: $F_A = \frac{s_A^2}{s_R^2}$; $F_B = \frac{s_B^2}{s_R^2}$.

Эти величины подчиняются F-распределению с числами степеней свободы, равными m_A, m_R и m_B, m_R соответственно. Здесь $m_A = I - 1$, $m_B = J - 1$, $m_R = m_A m_R$.

Зададим уровень значимости α и найдем в таблице квантилей F – распределения критические значения $F_{крит}$ для этих степеней свободы (для каждой гипотезы свое критическое значение, причем γ = 1 - α).

Правило

Если $F_A > F_{крит}$, то гипотеза H_A отвергается, если $F_A \le F_{крит}$, то гипотеза принимается.

Аналогичное правило для H_B.

Результаты дисперсионного анализа обычно сводят в таблицу 5.

Таблица 5. Дисперсионный анализ

Источник изменчивости	Сумма квадратов	Степени свободы	Средняя сумма квадратов	F-отношение
фактор A	S_A	$I-1$	s_A^2	$F_A = s_A^2 / s_R^2$

Окончание табл. 5

Источник изменчивости	Сумма квадратов	Степени свободы	Средняя сумма квадратов	F-отношение
фактор B	S_B	$J-1$	s_B^2	$F_B = s_B^2 / s_R^2$
ошибка эксперимента	S_R	$(I-1)(J-1)$	s_R^2	–
–	$S_П = S_A + S_B + S_R$	$(I-1)(J-1)-1$	–	–

__Пример__

Таблица 6. Исходные данные

B \ A	1	2	3	4	Средние по строке
1	2,68	3,29	2,88	4,45	3,32
2	4,12	4,96	5,09	5,22	4,85
3	5,52	4,50	5,42	5,29	5,18
Средние по столбцу	4,11	4,25	4,46	4,99	4,45

Таблица 7. Дисперсионный анализ

Источник изменчивости	Сумма квадратов	Степени свободы	Средняя сумма квадратов	F-отношение
фактор A	7,84	$3-1=2$	3,92	1,41
фактор B	1,34	$4-1=3$	0,446	0,16
ошибка эксперимента	16,68	6	2,78	–
–	21,04	5	–	–

Задаем уровень значимости $\alpha = 0,05$ и находим из таблиц $F_{крит}$ для каждой из гипотез.

Для гипотезы H_A: $F_{крит} = F(0,95;2;6) = 5,14$.

Для гипотезы H_B: $F_{крит} = F(0,95;3;6) = 4,76$.

Очевидно, в обоих случаях F- отношение меньше, чем $F_{крит}$ (см. таблицу 7). Следовательно, на уровне значимости 0,05 экспериментальные данные не противоречат этим гипотезам. Иначе говоря, следует принять гипотезы о том, что факторы A и B не влияют на переменную y.

5. Двухфакторный анализ с многократными наблюдениями

Как и в предыдущем случае, изучается влияние двух факторов на некоторую переменную. Отличие состоит в том, что для каждой комбинации уровней имеется более одного наблюдения. Ограничимся простейшим случаем двух наблюдений для каждой комбинации уровней. В этом случае таблица исходных данных содержит по два наблюдения в каждой ячейке. Большее количество наблюдений позволяет рассматривать более сложную модель, учитывающую возможное взаимодействие факторов. Модель имеет вид:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1,\dots,I;\ j = 1,\dots,J;\ k = 1,2.$$

Здесь μ – общее среднее значение; α_i, β_j – "чистые" (без учета взаимодействия) эффекты факторов A и B; γ_{ij} – эффекты взаимодействия факторов; ε_{ijk} – случайные ошибки (они удовлетворяют тем же условиям, что и в предыдущих пунктах).

Методом наименьших квадратов находят оценки параметров модели.

$$y_{ij\bullet} = \frac{1}{2}\left(y_{ij1} + y_{ij2}\right);\ y_{i\bullet\bullet} = \frac{1}{J}\sum_{j=1}^{J} y_{ij\bullet};\ y_{\bullet j\bullet} = \frac{1}{I}\sum_{i=1}^{I} y_{ij\bullet};\ y_{\bullet\bullet\bullet} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij\bullet}$$

$$\hat{\alpha}_i = y_{i\bullet\bullet} - y_{\bullet\bullet\bullet};\ \hat{\beta}_j = y_{\bullet j\bullet} - y_{\bullet\bullet\bullet};\ \hat{\gamma}_{ij} = y_{ij\bullet} - y_{\bullet\bullet\bullet}.$$

Далее для дисперсионного анализа нужно вычислить следующие суммы квадратов.

$$S_A = 2J\sum_{i=1}^{I}\hat{\alpha}_i^2;\quad S_B = 2I\sum_{j=1}^{J}\hat{\beta}_j^2;\quad S_{AB} = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\hat{\gamma}_{ij}^2;$$

$$S_R = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{2}\left(y_{ijk} - y_{ijg}\right)^2 .$$

Делим эти суммы на соответствующие числа степеней свободы.

$$s_A^2 = S_A/(I-1);\quad s_B^2 = S_B/(J-1);\quad s_{AB}^2 = S_{AB}/(I-1)/(J-1);$$

$$s_R^2 = S_R/(I-1)/(J-1).$$

Результаты дисперсионного анализа удобно представлять в виде следующей таблицы.

Таблица 8. Результаты дисперсионного анализа

Источник именчивости	Суммы квадратов	Степени свободы	Средние суммы квадратов	F-отношение
фактор А	S_A	$I-1$	s_A^2	$F_A = s_A^2 / s_R^2$
фактор В	S_B	$J-1$	s_B^2	$F_B = s_B^2 / s_R^2$
взаимодействие факторов	S_{AB}	$(I-1)(J-1)$	s_{AB}^2	$F_{AB} = s_{AB}^2 / s_R^2$
ошибка эксперимента	S_R	$(I-1)(J-1)$	s_R^2	–
–	S_{II}	$2IJ-1$	–	–

Проверяются три гипотезы:

$H_A: \alpha_i = 0,\ i = 1,\dots,I$ – фактор A не влияет на y;

$H_B: \beta_j = 0,\ j = 1,\dots,J$ – фактор B не влияет на y;

$H_{AB}: \gamma_{ij} = 0,\ i = 1,\dots,I,\ j = 1,\dots,J$ – взаимодействие факторов отсутствует.

Пример.

Таблица 9. Исходные данные $y_{ijk}, i = 1,2,3;\ j = 1,2,3,4;\ k = 1,2$.

A \ B	1	2	3	4
1	2,38; 2,68	2,61; 2,89	2,87; 3,03	2,82; 2,46
2	4,20; 4,12	3,96; 4,02	5,09; 4,85	4,78; 5,22
3	5,52; 5,98	4,15; 5,05	5,42; 5,64	5,29; 5,87

Таблица 10. Средние значения $y_{ijg} = \frac{1}{2}(y_{ij1} + y_{ij2})$

A \ B	1	2	3	4	Средние по строкам
1	2,53	2,40	2,55	2,64	2,53
2	4,16	4,64	4,97	5,00	4,69
3	5,75	4,90	5,53	5,58	5,44
Средние по столбцам	4,15	3,98	4,35	4,41	4,22

Таблица 11. Дисперсионный анализ

Источник именчивости	Суммы квадратов	Степени свободы	Средние суммы квадратов	F-отношение
фактор А	$S_A = 36,54$	2	$s_A^2 = 18,27$	$F_A = 93,89$
фактор В	$S_B = 0,69$	3	$s_B^2 = 0,23$	$F_B = 1,18$
взаимодействие факторов	$S_{AB} = 38,35$	6	$s_{AB}^2 = 6,39$	$F_{AB} = 32,84$
ошибка эксперимента	$S_R = 2,34$	12	$s_R^2 = 0,19$	–
–	$S_П = 75,58$	23	–	–

Находим из таблицы F - распределения критические значения при α = 0,05 и степенях свободы (2; 12), (3; 12) и (6; 12) и сравниваем их с рассчитанными.

Получаем:

гипотеза $H_A : F_{крит} = 3,89 < F_A = 93,89$ - гипотеза отвергается;

гипотеза $H_B : F_{крит} = 3,49 > F_B = 1,18$ - гипотеза принимается;

гипотеза $H_{AB} : F_{крит} = 3,00 < F_{AB} = 32,84$ - гипотеза отвергается.

Выводы:

- фактор A влияет на переменную y;
- фактор B при фиксированном уровне фактора A не влияет на переменную y;
- действие фактора B проявляется в том, что степень влияния фактора A зависит от того, на каком уровне находится фактор B.

Замечание.

В данном примере рассчитанные значения F значительно отличаются от критических, поэтому даже заметное изменение уровня значимости не отменит сделанных выводов. При других данных тот или иной выбор значения α может существенно повлиять на выводы. Предположим, например, что мы получили рассчитанное значение F_B, равное 3,18 вместо 1,18. Тогда, выбрав большее значение $\alpha = 0,1$, получим меньшее значение $F_{крит} = 2,61 < F_B$. В этом случае гипотезу H_B в соответствии с правилом пришлось бы отвергнуть.

Библиографический список

1. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1979.
2. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики (для технических приложений). – М.: Наука, 1969.
3. Планирование эксперимента в исследовании технологических процессов. – М.: Мир, 1977.

Приложение. Варианты индивидуальных заданий по теме: "Двухфакторный дисперсионный анализ с однократными наблюдениями".

вариант 1

	B1	B2	B3	B4
A1	0.96	1.04	0.40	-0.03
A2	0.87	-0.04	-0.46	0.11
A3	0.27	0.97	0.35	-0.81

вариант 2

	B1	B2	B3	B4
A1	2.68	3.29	2.88	4.45
A2	4.12	4.96	5.09	5.22
A3	5.52	4.50	5.42	5.29

вариант 3

	B1	B2	B3	B4
A1	1.08	2.52	1.75	2.65
A2	1.80	2.42	2.13	3.78
A3	1.83	2.42	2.65	3.00

вариант 4

	B1	B2	B3	B4
A1	1.80	2.14	1.63	0.86
A2	3.96	1.81	1.41	2.05
A3	5.00	4.48	2.61	3.19

вариант 5

	B1	B2	B3	B4
A1	3.11	1.61	2.72	2.68
A2	3.40	2.76	3.21	2.82
A3	5.20	4.36	4.97	5.60

вариант 6

	B1	B2	B3	B4
A1	1.90	1.59	0.98	1.66
A2	4.22	2.21	2.77	1.45
A3	4.30	4.16	4.11	2.60

вариант 7

	B1	B2	B3	B4
A1	0.71	1.23	0.22	0.30
A2	3.53	1.71	2.72	2.17
A3	2.97	3.88	2.26	3.40

вариант 8

	B1	B2	B3	B4
A1	1.68	1.86	1.57	2.17
A2	1.70	2.92	2.10	2.07
A3	2.34	2.93	3.94	4.22

вариант 9

	B1	B2	B3	B4
A1	2.91	1.87	2.23	2.43
A2	4.16	3.63	5.08	5.11
A3	4.38	4.62	5.01	6.00

вариант 10

	B1	B2	B3	B4
A1	0.71	1.32	2.14	2.81
A2	3.01	2.51	2.16	2.77
A3	3.52	2.24	3.37	4.14

вариант 11

	B1	B2	B3	B4
A1	3.60	3.62	3.51	3.33
A2	3.90	5.41	5.18	4.56
A3	6.38	6.21	7.01	7.72

вариант 12

	B1	B2	B3	B4
A1	1.52	2.42	2.19	2.51
A2	2.66	2.73	3.68	4.99
A3	4.18	5.54	4.56	4.87

вариант 13

	B1	B2	B3	B4
A1	2.36	2.70	3.41	2.75
A2	4.04	4.40	4.81	3.46
A3	4.83	4.61	6.07	6.40

вариант 14

	B1	B2	B3	B4
A1	0.47	2.17	1.51	1.93
A2	1.44	1.55	2.06	2.22
A3	2.96	3.03	2.55	2.75

вариант 15

	B1	B2	B3	B4
A1	1.84	2.24	0.64	1.50
A2	3.04	2.25	1.91	1.21
A3	4.78	3.81	4.03	3.33

вариант 16

	B1	B2	B3	B4
A1	1.07	2.07	1.00	-0.71
A2	1.54	2.85	2.15	1.28
A3	3.28	3.04	1.72	1.12

вариант 17

	B1	B2	B3	B4
A1	2.78	3.29	3.33	3.13
A2	3.35	3.94	3.48	4.15
A3	3.04	5.13	4.28	5.46

вариант 18

	B1	B2	B3	B4
A1	3.18	2.75	3.49	3.36
A2	4.08	4.48	4.71	5.16
A3	5.10	6.06	6.20	5.75

вариант 19

	B1	B2	B3	B4
A1	1.86	0.93	0.66	0.65
A2	3.27	2.37	2.55	1.80
A3	3.82	3.40	4.47	4.42

вариант 20

	B1	B2	B3	B4
A1	1.01	0.30	1.09	0.79
A2	1.25	0.68	1.64	-0.09
A3	0.58	0.46	0.46	0.03

вариант 21

	B1	B2	B3	B4
A1	1.47	1.40	0.77	0.36
A2	2.96	2.05	2.76	2.42
A3	2.08	3.15	3.35	3.34

вариант 22

	B1	B2	B3	B4
A1	1.60	1.47	0.36	1.32
A2	4.18	3.59	3.95	2.44
A3	6.31	5.26	5.41	4.59

вариант 23

	B1	B2	B3	B4
A1	3.13	1.81	3.71	2.11
A2	2.86	3.63	4.72	3.25
A3	4.44	4.73	4.78	5.61

вариант 24

	B1	B2	B3	B4
A1	2.46	2.18	2.43	1.33
A2	3.17	4.25	2.67	2.07
A3	4.45	5.48	4.08	3.66

Содержание

Абжандадзе Зураб Леванович
Федорова Евгения Владимировна
Юдовин Марк Эльявич

ЭЛЕМЕНТЫ ДИСПЕРСИОННОГО АНАЛИЗА

Методические указания

для студентов всех специальностей дневной формы обучения

Редактор и корректор В.А. Басова
Техн. редактор Л.Я. Титова

Подп. к печати 09.06.09 г. Формат 60×84/16. Бумага тип № 1.
Печать офсетная. Объём 1,5 печ.л., 1,5 уч.-изд. л.
Тираж 50 экз. Изд. № 40. Цена «С». Заказ 2014.

Ризограф ГОУВПО Санкт-Петербургского государственного технологического университета растительных полимеров, 198095, Санкт-Петербург, ул. Ивана Черных, 4.