С.Ж.АСФЕНДИЯРОВ АТЫНДАҒЫ
ҚАЗАҚ ҰЛТТЫҚ МЕДИЦИНА
УНИВЕРСИТЕТІ

КАЗАХСКИЙ НАЦИОНАЛЬНЫЙ МЕДИЦИНСКИЙ
УНИВЕРСИТЕТ ИМЕНИ С.Д.АСФЕНДИЯРОВА

КАФЕДРА МЕДИЦИНСКОЙ БИОФИЗИКИ, ИНФОРМАТИКИ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

СРС на тему:

Двухфакторный дисперсионный анализ.

Выполнила: студентка 3 курса, факультета ОМ, 58 группы,
Зосимова Анастасия
Проверила: Исмаилова Мадина Маликовна

Алматы,2011

ПЛАН:

1. ВВЕДЕНИЕ
2. Назначение
3. Обработка двухфакторного дисперсионного комплекса
4. Схема двухфакторного дисперсионного анализа
5. Разновидности метода

 -есть повторные измерения

 -нет повторных измерений

6. Пример
7. Заключение
8. Литература

ВВЕДЕНИЕ

Техника дисперсионного анализа полезна для ряда статистических задач, связанных с исследованием влияния одной или нескольких качественных переменных (*факторов*) на одну зависимую количественную переменную (*отклик*).

- **Общая дисперсия:**

$$C_0 = \sum (X - X_C)^2$$

- **Факториальная дисперсия:**

$$C_\phi = \sum (X_\phi - X_0)^2$$

- **Случайная дисперсия:**

$$C_c = \sum (X - X_\phi)^2$$

- **Где X – отдельное значение результативного признака**
- **X_c – общая средняя арифметическая всего комплекса**
- **$X_ф$ – групповая средняя**

Назначение.

Посредством данного метода в зависимости от типа модели по каждому фактору (с фиксированными или же со случайными эффектами) с помощью параметрического критерия Фишера проверяется одна из двух нулевых гипотез:

- средние значения для групп откликов, измеренных при различных значениях фактора, не имеют существенных различий между собой;
- дисперсия средних значений для групп откликов, измеренных при различных значениях фактора, не отлична от нуля.

Обработка двухфакторного дисперсионного комплекса

1. Вычисление общей дисперсии осуществляется как при однофакторном комплексе
2. Вычисление случайной дисперсии аналогично нахождению ее в однофакторном комплексе
3. Вычисление дисперсии суммарного действия организованных факторов

Схема двухфакторного дисперсионного анализа

Источник вариации	Сумма квадратов отклонений D	Число степеней. свободы d. f.	Средний квадрат отклонений s2 = D/d.f.	F-критерий
Факторы x и z	D′факт•K	mp - 1	s2факт	
Фактор x	D′x•K	m - 1	s2x	$F = \frac{s_x^2}{s_{ост}^2}$
Фактор z	D′z•K	p - 1	s2z	$F = \frac{s_z^2}{s_{ост}^2}$
Взаимодействие факторов x и z	(D′факт- D′x- - D′z)•K	mp – p-m+1	s2xz	$F = \frac{s_{xz}^2}{s_{ост}^2}$
Остаточная	Dобщ - D′факт•K	n - mp	s2ост	
Общая	Dобщ	n - 1	s2	

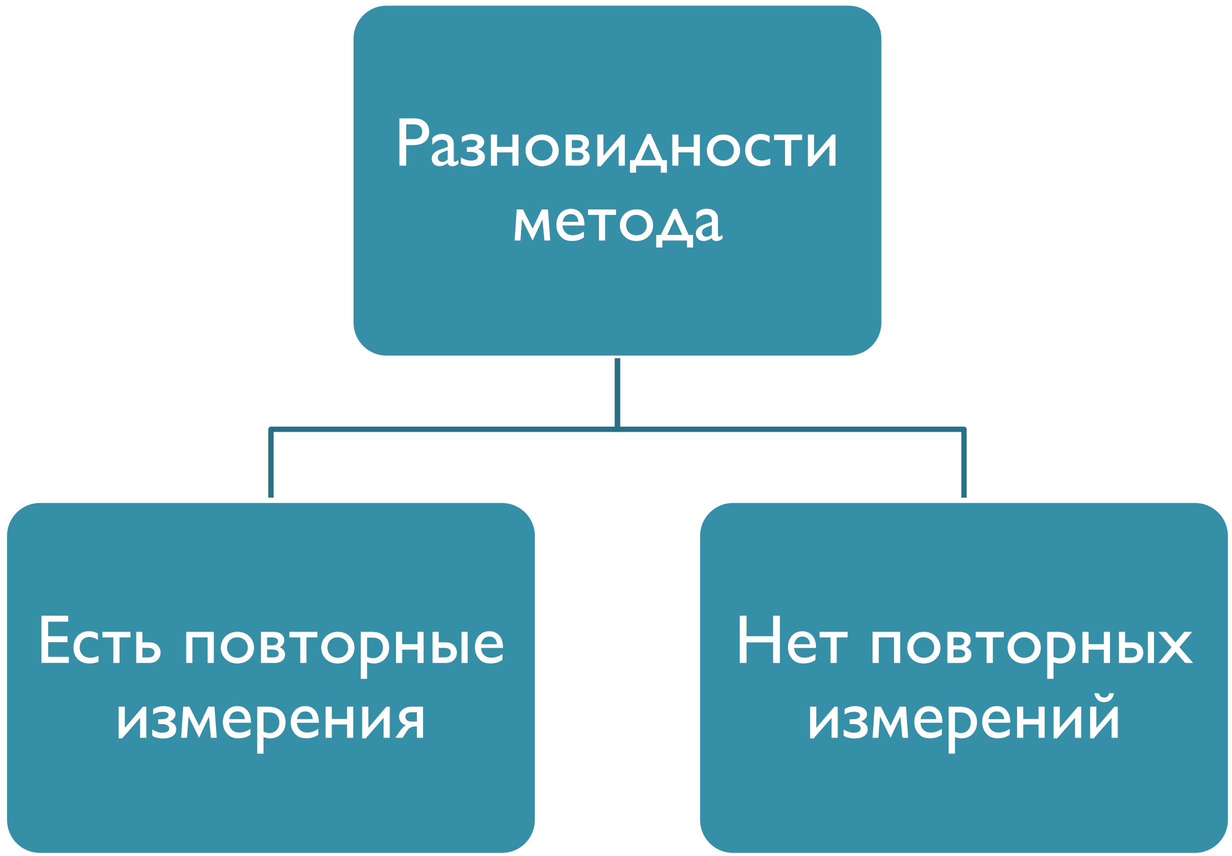
В двухфакторном дисперсионном анализе испытуемые гипотезы формулируются следующим образом:

1. $H0 : \mu 1\bullet = \mu 2. = \ldots \mu m$

2. $H0 : \mu 1\bullet = \mu 2. = \ldots \mu p$

3. $H0 : \mu 1\bullet = \mu 2. = \ldots \mu mp$

Имеется две разновидности метода в зависимости от того, производились ли *повторные измерения* при каждом сочетании двух исследуемых факторов или нет.



Нет повторных измерений.

При эксперименте без повторных измерений исходные данные должны представлять собой матрицу размером $m \cdot n$, в которой столбцы отвечают различным уровням первого фактора $j=1,...,m$, строки отвечают различным уровням второго фактора $i=1,...,n$, а каждая ячейка содержит отклик измеренный при соответствующем сочетании уровней исследуемых факторов.

Выдача: выдача включает дисперсионную таблицу со столбцами: сумма квадратов, число степеней свободы, средняя сумма квадратов, сила влияния фактора (по Снедекору), а строки содержат значения для первого и второго факторов, а так же остаточные и общие параметры.

Далее для каждого фактора вычисляется статистика Фишера F с уровнем значимости P. Если $P>0.05$, нулевая гипотеза об отсутствии влияния фактора может быть принята.

Есть повторные измерения.

При эксперименте с повторными измерениями исходные данные должны представлять собой псевдоматрицу (не обязательно одинаковой длинны столбцов), в которой переменные (i=1,..., m$\cdot n$) отвечают различным уровням исследуемых факторов в порядке изменения значений первого фактора: все уровни первого фактора для первого уровня второго фактора, все уровни первого фактора для второго уровня второго фактора и т.д., а каждая переменная содержит J_i откликов (J_i >1), измеренных при данном сочетании значений факторов.

Есть повторные измерения.

Выдача: выдача включает дисперсионную таблицу со столбцами: сумма квадратов, число степеней свободы, средняя сумма квадратов, сила влияния фактора (по Снедекору), а строки содержат значения для первого и для второго факторов, для эффекта межфакторного взаимодействия, а так же остаточные и общие параметры.

Далее для каждого фактора вычисляется статистика Фишера F с уровнем значимости P. Если $P>0.05$, нулевая гипотеза об отсутствии влияния фактора может быть принята.

Если эффект взаимодействия не обнаружен, то проводится дополнительный анализ по факторам A и B, но без учета их взаимодействия. Такой дополнительный анализ, как правило, дает более низкий уровень значимости нулевых гипотез. Полученными результатами рекомендуется пользоваться, если уровень значимости гипотезы отсутствия взаимодействия факторов достаточно велик ($P>0.05$).

Бесповторный эксперимент.

Формулы.

В случае двухфакторного эксперимента без повторных измерений дисперсионная таблица имеет вид:

Источник	Сумма Квадратов	Степени свободы	Средн. квадр.	Сила влияния
Фактор 1	$SA = n \cdot \sum_i (X_i - X)^2$	m-1	A=SA/(m-1)	h_a
Фактор 2	$SB = k \cdot \sum_i (X_i - X)^2$	n-1	B=SB/(n-1)	h_b
Остаточная	$SE = \sum_{ij} (x_{ij} - X_i - X_j + X)^2$	(m-1) · (n-1)	$E = \frac{SE}{(n-1)\cdot(m-1)}$	
Общее	$T = \sum_{ij} (x_{ij} - x)^2$	m+n-1		

где:

$$X_i = \sum_j x_{ij} \qquad X_j = \sum_i x_{ij} \qquad X = \sum_{ij} x_{ij} \qquad h_a = \frac{SA - SE}{SA + SE} \qquad h_a = \frac{SB - SE}{SB + SE}$$

$$F_1 = \frac{A}{E}, F_2 = \frac{B}{E}$$

Повторы и фиксированные эффекты.

В случае двухфакторного эксперимента с повторными измерениями и с фиксированными эффектами дисперсионная таблица имеет вид:

Источник	Сумма Квадратов	Степени свободы	Средн. квадр.	Сила влияния
Фактор 1	SA	m-1	A=SA/(m-1)	h_a
Фактор 2	SB	n-1	B=SB/(n-1)	h_b
Мефактор.	$SAB = (SX_i - SA - SB)\cdot K$	$(m-1)\cdot(n-1)$	$AB = \frac{SAB}{(n-1)\cdot(m-1)}$	h_{ab}
Остаточная	$SE = \sum_{ij} x_{ij}^2 - \frac{1}{N}\cdot\sum X_i^2$	$N-m\cdot n$	$E = \frac{SE}{N-n\cdot m}$	
Общее	$T = \sum_{ij} x_{ij}^2 - H$	N-1		

где:

$$A = N\cdot\left(\sum_a\left(\frac{X_a}{n}\right)^2 - H_1\right) \qquad B = N\cdot\left(\sum_b\left(\frac{X_b}{m}\right)^2 - H_1\right) \qquad K = \frac{SX}{SX_1} \qquad SX = \sum_i \frac{X_i^2}{n_i} - H$$

$$SX_1 = N\cdot\left(\frac{1}{n\cdot m}\sum_i \frac{X_i^2}{n_i^2} - H_1\right) \qquad H = \frac{1}{N}\cdot\left(\sum_{ij} x_{ij}\right)^2 \qquad H_1 = \frac{1}{(m\cdot n)^2}\cdot\left(\sum_i X_i\right)^2$$

X_i

X_a X_b

$$h_a = \frac{SA-SE}{SA+SE} \qquad h_a = \frac{SB-SE}{SB+SE} \qquad h_{ab} = \frac{SAB-SE}{SAB+SE}$$

F - статистики с n-1, $n\cdot m\cdot(k-1)$; m-1, $n\cdot m\cdot(k-1)$; $(n-1)\cdot(m-1)$, $n\cdot m\cdot(k-1)$ степенями свободы, k=N/n

Примечания.

- Отличие модели со случайными эффектами состоит в замене второго числа степеней свободы в F_1 F_2 - статистиках (n-1)·(m-1);
- Отличие модели с рандомизованными блоками состоит в замене второго числа степеней свободы в F_1 - статистиках (n-1)·(m-1);
- Отличие модели с группировкой - вычисляются два F - значения:

F_1=A/B с n-1, n·m·(k-1) степенями свободы;

F_2 =B/E с n·(m-1), n·m·(k-1) степенями свободы, вычисление межфакторного взаимодействия не производится;

- В случае незначительного межфакторного взаимодействия при повторных вычислениях F_1 F_2

используется E=E+AB с (n-1)·(m-1)+N-n·m степенями свободы.

Ниже будет проанализирован наиболее простой случай с двумя факторами, воздействующими на результат эксперимента. При этом мы ограничимся факторами F и G с фиксированными уровнями F_1, $F_2, \ldots, F_p$ и G_1, $G_2, \ldots, G_q$ соответственно и, кроме того, будем предполагать, что взаимодействие между факторами отсутствует и что при каждой комбинации уровней F_i, G_j, производится лишь по одному наблюдению x_{ij}. Модель такого эксперимента можно записать в виде

$$x_{ij} = F_i + G_j + \varepsilon_{ij}, \quad 1 \leqslant i \leqslant p, \quad 1 \leqslant j \leqslant q,$$

где ε_{ij} — независимые нормально распределенные случайные ошибки с нулевым средним и общей дисперсией σ^2.

Результаты наблюдений удобно представить в виде таблицы:

Фактор G	Фактор F				
	F_1	F_2	...	F_p	Суммы по строкам
G_1	x_{11}	x_{21}	...	x_{p1}	$x_{\cdot 1}$
G_2	x_{12}	x_{22}	...	x_{p2}	$x_{\cdot 2}$
...	...	...	...	...	...
G_q	x_{1q}	x_{2q}	...	x_{pq}	$x_{\cdot q}$
Суммы по столбцам	$x_{1\cdot}$	$x_{2\cdot}$	...	$x_{p\cdot}$	$x_{\cdot\cdot}$

Здесь приняты обозначения $x_{\cdot j}$ $(j = 1, 2, \ldots, q)$ для сумм наблюдений по строкам и $x_{i\cdot}$ $(i = 1, 2, \ldots, p)$ для сумм наблюдений по столбцам, $x_{\cdot\cdot}$ — общая сумма всех наблюдений.

Факторные S_F^2, S_G^2 и остаточную $S_{\text{ост}}^2$ дисперсии можно вычислить по формулам

$$S_F^2 = \frac{1}{p-1}\left(\frac{1}{q}\sum_{i=1}^{p}(x_{i\cdot} - \beta)^2 - \frac{1}{pq}(x_{\cdot\cdot} - p\beta)^2\right),$$

$$S_G^2 = \frac{1}{q-1}\left(\frac{1}{p}\sum_{i=1}^{q}(x_{\cdot j} - \beta)^2 - \frac{1}{pq}(x_{\cdot\cdot} - q\beta)^2\right),$$

$$S_{\text{ост}}^2 = \frac{1}{(p-1)(q-1)}\left(S^2 - (p-1)S_F^2 - (q-1)S_G^2\right),$$

где

$$S^2 = \sum_{i,j}(x_{ij} - \beta)^2 - \frac{1}{pq}(x_{\cdot\cdot} - pq\beta)^2,$$

β — произвольное число, которое может быть разным в различных формулах.

При этом (см. (8.96)) $M(S_{\text{ост}}^2) = \sigma^2$, т. е. $S_{\text{ост}}^2$ является несмещенной оценкой дисперсии σ^2. Аналогичным свойством, в предположениях H_F: $F_1 = F_2 = \ldots = F_p$ и H_G: $G_1 = G_2 = \ldots = G_q$ соответственно, обладают факторные дисперсии S_F^2 и S_G^2.

Таким образом, так же как в п. 8.6.1, мы можем составить отношения $\Lambda_F = \dfrac{S_F^2}{S_{\text{ост}}^2}$ или $\Lambda_G = \dfrac{S_G^2}{S_{\text{ост}}^2}$ и с их помощью проверить нулевые гипотезы H_F или H_G, сравнивая Λ_F или Λ_G с критической точкой $f_{\text{кр}}(\alpha; k_1, k_2)$ F- распределения (приложение 3) со степенями свободы $k_1 = q - 1$ или $p - 1$, и $k_2 = (p-1)(q-1)$.

Например, чтобы проверить существенность влияния фактора F на результаты эксперимента, следует, задавшись уровнем значимости α, сравнить наблюдаемое значение $\Lambda_{F,\text{набл}}$ статистики критерия Λ_F с величиной $f_{\text{кр}} = f_{\text{кр}}(\alpha; p-1, (p-1)(q-1))$ из приложения 3.

Пример В условиях модели определить по данным табл. достоверность влияния препаратов F и G на массу подопытных животных при уровне значимости $\alpha = 0{,}05$.

Таблица

Фактор G	Фактор F			
	F_1	F_2	F_3	Суммы по строкам
G_1	30	35	40	105
G_2	32	39	38	109
G_3	34	38	44	116
G_4	28	36	42	106
Суммы по столбцам	124	148	164	436

Решение. Сначала найдем S_F^2. Имеем при $p = 3$, $q = 4$ и $\beta = 148$)

$$S_F^2 = \frac{1}{2}\left(\frac{1}{4}(-24)^2 + 0^2 + 16^2) - \frac{1}{12}(436 - 3 \cdot 148)^2\right) = 101{,}33.$$

Аналогично при $\beta = 109$

$$S_G^2 = \frac{1}{3}\left(\frac{1}{3}(-4)^2 + 0^2 + 7^2 + (-3)^2) - \frac{1}{12}(436 - 4 \cdot 109)^2\right) = 8{,}22$$

и (при $\beta = 38$)

$$S^2 = (-8)^2 + (-3)^2 + 2^2 + (-6)^2 + 1^2 + 0^2 + (-4)^2 + 0^2 + 6^2 + \\ + (-10)^2 + (-2)^2 + 4^2 - \frac{1}{12}(436 - 12 \cdot 38)^2 = 252{,}67,$$

$$S^2_{\text{ост}} = \frac{1}{2 \cdot 3}(252{,}67 - 2 \cdot 101{,}33 - 3 \cdot 8{,}22) = \frac{25{,}35}{6} = 4{,}22.$$

Отсюда $\Lambda_{F,\text{набл}} = 24{,}01$ и $\Lambda_{G,\text{набл}} = 1{,}95$. Учитывая что $f_{\text{кр}}(005; 3, 6) = 4{,}76 > 1{,}95$, а $f_{\text{кр}}(005; 2, 6) = 5{,}14 < 24{,}01$, делаем вывод, что препарат F оказывает влияние на массу подопытных животных, а препарат G такого влияния не оказывает.

Если $\Lambda_{F,\text{набл}} < f_{\text{кр}}$, то нулевая гипотеза H_F об отсутствии влияния фактора F на результаты эксперимента подтверждается, если же $\Lambda_{\text{набл}} \geqslant f_{\text{кр}}$, то гипотеза H_F отвергается на уровне α и делается вывод, что фактор F воздействует на результаты.

Заметим, что если $\Lambda_{\text{ост}} \leqslant 1$, то нулевая гипотеза принимается сразу.

Пример.
Матрица, четыре переменные в которой представляют результаты побед в четырех видах спорта (плавание, борьба, прыжки в высоту, шахматы - 1-й, 2-й, 3-й, 4-й столбцы соответственно). Необходимо выяснить, влияет ли вес и рост спортсменов на их спортивные достижения. Замеры веса и роста проводились через равные промежутки времени у спортсменов примерно одинаковой квалификации, но с разными показателями роста и веса.
Исходные данные:

Плавание	Борьба	Прыжки в высоту	Шахматы
6	12	9	10
8	11	7	9
10	8	9	9
11	12	15	12
10	9	12	8

Результаты:

Источник	Сум.квадр	Ст.своб	Ср.квадр	Сила влияния
Факт.1	6,95	3	2,32	
Факт.2	37,3	4	9,33	
Остат.	40,3	12	3,36	
Общая	84,6	19	4,45	

2-ФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ.
Факторный план: неповторяемый
F(фактор1)=0,69, Значимость=0,578, степ.своб = 3,12
Гипотеза 0: <Нет влияния фактора на отклик>
F(фактор2)=2,78, Значимость=0,0758, степ.своб = 4,12
Гипотеза 0: <Нет влияния фактора на отклик>
Параметры модели:
Среднее = 9,85, доверит.инт.=3,48
Эффект1-1 = -0,85, доверит.инт.=28
Эффект1-2 = 0,55, доверит.инт.=28
Эффект1-3 = 0,55, доверит.инт.=28
Эффект1-4 = -0,25, доверит.инт.=28
Эффект2-1 = -0,6, доверит.инт.=33,5
Эффект2-2 = -1,1, доверит.инт.=33,5
Эффект2-3 = -0,85, доверит.инт.=33,5
Эффект2-4 = 2,65, доверит.инт.=33,5
Эффект2-5 = -0,1, доверит.инт.=33,5
Вывод: Дисперсионный анализ не обнаруживает существенного влияния роста и веса спортсменов на количество побед в соревнованиях.

ЗАКЛЮЧЕНИЕ

Следует отметить, что принципиальной разницы между двухфакторным и однофакторным дисперсионным анализом нет. Двухфакторный анализ не меняет общую логику дисперсионного анализа, а лишь несколько усложняет ее, поскольку, кроме учета влияния на зависимую переменную каждого из факторов по отдельности, следует оценивать и их совместное действие. Таким образом, то новое, что вносит в анализ данных двухфакторный дисперсионный анализ, касается в основном возможности оценить межфакторное взаимодействие. Тем не менее, по-прежнему остается возможность оценивать влияние каждого фактора в отдельности. В этом смысле процедура двухфакторного дисперсионного анализа более экономична, поскольку решает сразу две задачи: оценивается влияние каждого из факторов и их взаимодействие.

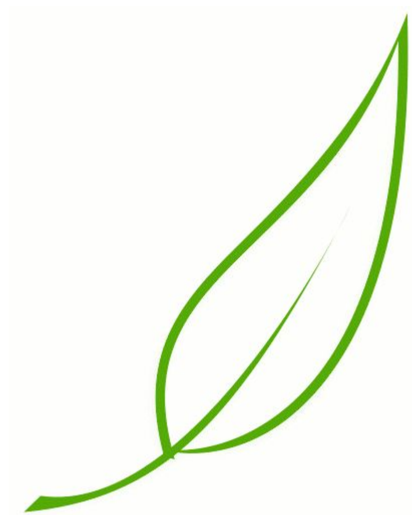
2'

'

Ф

'''

''''



''

3

ЛИТЕРАТУРА:

1.В. М.Зайцев, В. Г. Лифляндский, В. И. Маринкин ПРИКЛАДНАЯ МЕДИЦИНСКАЯ СТАТИСТИКА,2003г.
2.И.В. Павлушков Основы высшей математики и математической статистики. (учебник для медицинских и фармацевтических вузов), 2008г.
3.Гланц С. Медико-биологическая статистика – М.:Практика,1999.
4.http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content353/Content353.htm