Дисперсійний аналіз даних ANOVA analysis of variance

ГРАФОВІ ЙМОВІРНІСНІ МОДЕЛІ

Сумський державний університет

Дисперсійний аналіз (англ. *analysis of variance ANOVA*)

являє собою статистичний метод аналізу результатів, які залежать від якісних ознак.

-

Дисперсійний аналіз (англ. *analysis of variance ANOVA*)

являє собою статистичний метод аналізу результатів, які залежать від якісних ознак.

-

Кожен фактор може бути дискретною чи неперервною випадковою змінною, яку розділяють на декілька сталих рівнів (градацій, інтервалів).

-

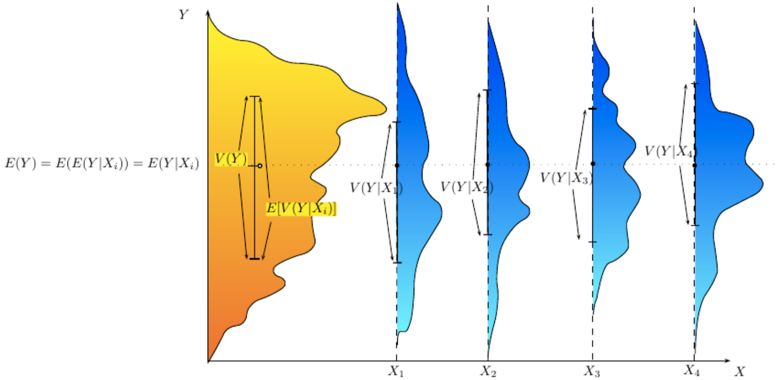


Figure 2: ANOVA : No fit

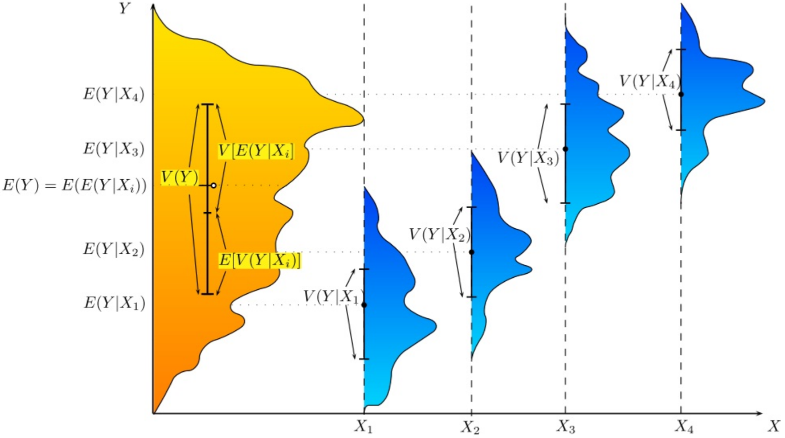


Figure 1: ANOVA : Fair fit

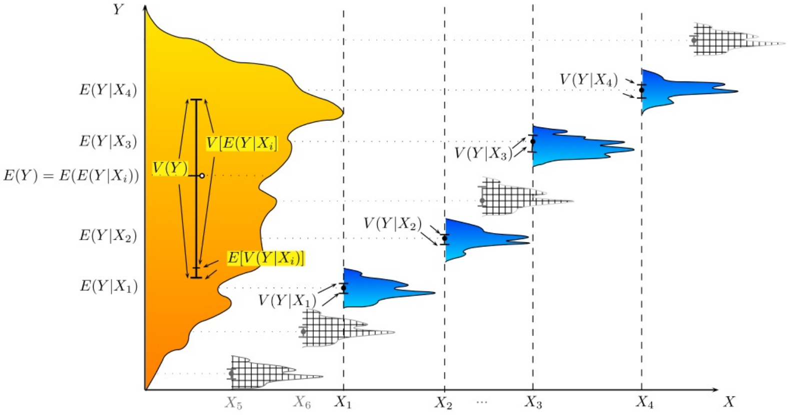


Figure 3: ANOVA : very good fit

В будь-якому експерименті середні значення досліджуваних величин змінюються у зв'язку зі зміною основних факторів (кількісних та якісних), що визначають умови досліду, а також і випадкових факторів. Дослідження впливу тих чи інших факторів на мінливість середніх є задачею дисперсійного аналізу.

-

В будь-якому експерименті середні значення досліджуваних величин змінюються у зв'язку зі зміною основних факторів (кількісних та якісних), що визначають умови досліду, а також і випадкових факторів. Дослідження впливу тих чи інших факторів на мінливість середніх є задачею дисперсійного аналізу.

-

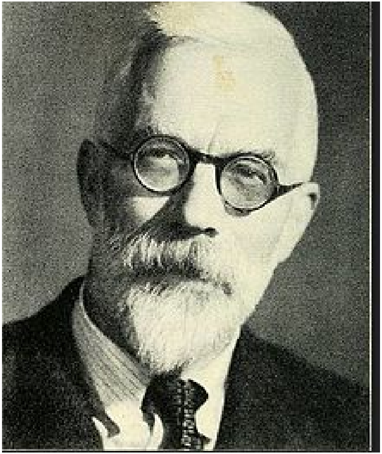
Дисперсійний аналіз використовує властивість адитивності дисперсії випадкової величини, що обумовлено дією незалежних факторів. В залежності від числа джерел дисперсії розрізняють однофакторний та багатофакторний дисперсійний аналіз.

-

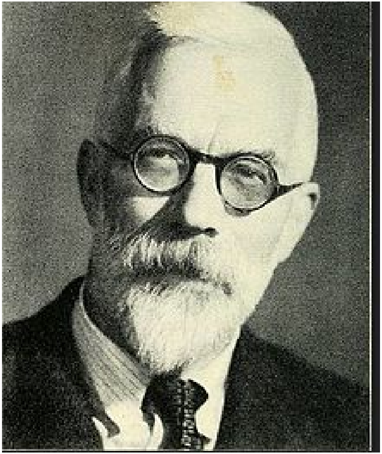
Дисперсійний аналіз полягає у виділенні і оцінці окремих факторів, що викликають зміну досліджуваної випадкової величини.
При цьому проводиться розклад сумарної вибіркової дисперсії на складові, обумовлені незалежними факторами.
Кожна з цих складових є оцінкою дисперсії генеральної сукупності.
Щоб вирішити, чи дієвий вплив даного фактору, необхідно оцінити значимість відповідної вибіркової дисперсії у порівнянні з дисперсією відтворення, обумовленою випадковими факторами.
Перевірка значимості оцінок дисперсії проводять по критерію Фішера.

Коли розрахункове значення критерію Фішера виявиться меншим табличного, то вплив досліджуваного фактору немає підстав вважати значимим. Коли ж розрахункове значення критерію Фішера виявиться більшим табличного, то цей фактор впливає на зміни середніх.

Методи дисперсійного аналізу були розроблені сером Рональдом Фішером (1890-1962), професором генетики Каліфорнійського університету, одним із провідних світових статистиків і біологів свого часу.



Методи дисперсійного аналізу були розроблені сером Рональдом Фішером (1890-1962), професором генетики Каліфорнійського університету, одним із провідних світових статистиків і біологів свого часу.



Він був першим, хто запропонував використовувати латинські та греко-латинські квадрати для дисперсійного аналізу даних.

Латинський квадрат n-го порядку – це таблиця розміру $n \times n$, заповнена n елементами множини M таким чином, що в кожному рядку і в кожному стовпці таблиці кожен елемент зустрічається в точності один раз

Латинський квадрат n-го порядку – це таблиця розміру $n \times n$, заповнена n елементами множини M таким чином, що в кожному рядку і в кожному стовпці таблиці кожен елемент зустрічається в точності один раз

Леонард Ейлер (1707 – 1783)


MELENCOLIA I

"Меланхолія" – різцева гравюра на міді німецького художника Альбрехта Дюрера, закінчена в 1514 році.

Дюрер склав перший в європейському мистецтві магічний квадрат, 4x4. Сума чисел в будь-якому рядку, стовпці, діагоналі, а також в кожній чверті (в тому числі в центральному квадраті) і сума кутових чисел дорівнює 34. Два середніх числа в нижньому ряду вказують дату створення картини (1514). Два крайніх числа в нижньому ряду відповідають ініціалами художника. У середніх квадратах першого стовпчика внесені виправлення – цифри деформовані.

a_1	a_2
a_2	a_1

a_1	a_2	a_3
a_2	a_3	a_1
a_3	a_1	a_2

a_1	a_2	a_3
a_3	a_1	a_2
a_2	a_3	a_1

a_1	a_2	a_3	a_4
a_2	a_3	a_4	a_1
a_3	a_4	a_1	a_2
a_4	a_1	a_2	a_3

a_1	a_2	a_3	a_4
a_4	a_1	a_2	a_3
a_3	a_4	a_1	a_2
a_2	a_3	a_4	a_1

a_1	a_2	a_3	a_4	a_5
a_2	a_3	a_4	a_5	a_1
a_3	a_4	a_5	a_1	a_2
a_4	a_5	a_1	a_2	a_3
a_5	a_1	a_2	a_3	a_4

a_1	a_2	a_3	a_4	a_5
a_3	a_4	a_5	a_1	a_2
a_5	a_1	a_2	a_3	a_4
a_2	a_3	a_4	a_5	a_1
a_4	a_5	a_1	a_2	a_3

a_1	a_2	a_3	a_4	a_5
a_4	a_5	a_1	a_2	a_3
a_2	a_3	a_4	a_5	a_1
a_5	a_1	a_2	a_3	a_4
a_3	a_4	a_5	a_1	a_2

a_1	a_2	a_3	a_4	a_5
a_5	a_1	a_2	a_3	a_4
a_4	a_5	a_1	a_2	a_3
a_3	a_4	a_5	a_1	a_2
a_2	a_3	a_4	a_5	a_1

a_1	a_2	a_3	a_4	a_5	a_6
a_2	a_3	a_4	a_5	a_6	a_1
a_3	a_4	a_5	a_6	a_1	a_2
a_4	a_5	a_6	a_1	a_2	a_3
a_5	a_6	a_1	a_2	a_3	a_4
a_6	a_1	a_2	a_3	a_4	a_5

a_1	a_2	a_3	a_4	a_5	a_6
a_6	a_1	a_2	a_3	a_4	a_5
a_5	a_6	a_1	a_2	a_3	a_4
a_4	a_5	a_6	a_1	a_2	a_3
a_3	a_4	a_5	a_6	a_1	a_2
a_2	a_3	a_4	a_5	a_6	a_1

0	1	2	3	4	5
1	0	3	2	5	4
2	3	4	5	0	1
3	2	5	4	1	0
4	5	0	1	2	3
5	4	1	0	3	2

♠	♥
♥	♠

A	B	C
C	A	B
B	C	A

Вітраж з латинським квадратом 7-го порядку в одному з коледжів Кембриджу, присвячений Р.Фішеру

0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
3	4	0	1	2
4	0	1	2	3

0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
3	4	0	1	2
4	0	1	2	3

0	1	2	3	4
3	4	0	1	2
1	2	3	4	0
4	0	1	2	3
2	3	4	0	1

0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
3	4	0	1	2
4	0	1	2	3

0	1	2	3	4
2	3	4	0	1
4	0	1	2	3
1	2	3	4	0
3	4	0	1	2

0	1	2	3	4
3	4	0	1	2
1	2	3	4	0
4	0	1	2	3
2	3	4	0	1

0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
3	4	0	1	2
4	0	1	2	3

0	1	2	3	4
2	3	4	0	1
4	0	1	2	3
1	2	3	4	0
3	4	0	1	2

0	1	2	3	4
3	4	0	1	2
1	2	3	4	0
4	0	1	2	3
2	3	4	0	1

0	1	2	3	4
4	0	1	2	3
3	4	0	1	2
2	3	4	0	1
1	2	3	4	0

a	b	c	d
b	a	d	c
c	d	a	b
d	c	b	a

α	β	γ	δ
γ	δ	α	β
δ	γ	β	α
β	α	δ	γ

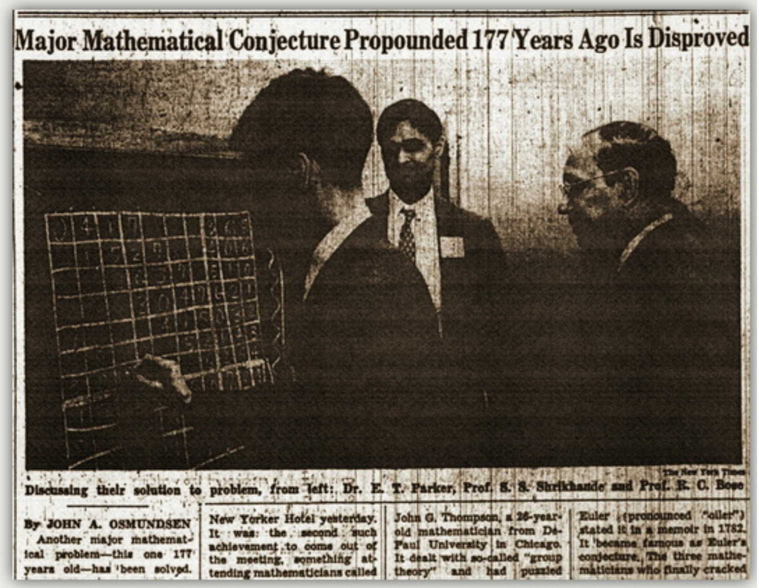
$a\ \alpha$	$b\ \beta$	$c\ \gamma$	$d\ \delta$
$b\ \gamma$	$a\ \delta$	$d\ \alpha$	$c\ \beta$
$c\ \delta$	$d\ \gamma$	$a\ \beta$	$b\ \alpha$
$d\ \beta$	$c\ \alpha$	$b\ \delta$	$a\ \gamma$


JESTER
STANDARD
PLAYING CARDS BY GPN

$$n = 4k + 2$$

00	47	18	76	29	93	85	34	61	52
86	11	57	28	70	39	94	45	02	63
95	80	22	67	38	71	49	56	13	04
59	96	81	33	07	48	72	60	24	15
73	69	90	82	44	17	58	01	35	26
68	74	09	91	83	55	27	12	46	30
37	08	75	19	92	84	66	23	50	41
14	25	36	40	51	62	03	77	88	99
21	32	43	54	65	06	10	89	97	78
42	53	64	05	16	20	31	98	79	87

The New-York Times



Major Mathematical Conjecture Propounded 177 Years Ago Is Disproved

The New York Times

Discussing their solution to problem, from left: Dr. E. T. Parker, Prof. S. S. Shrikhande and Prof. R. C. Bose

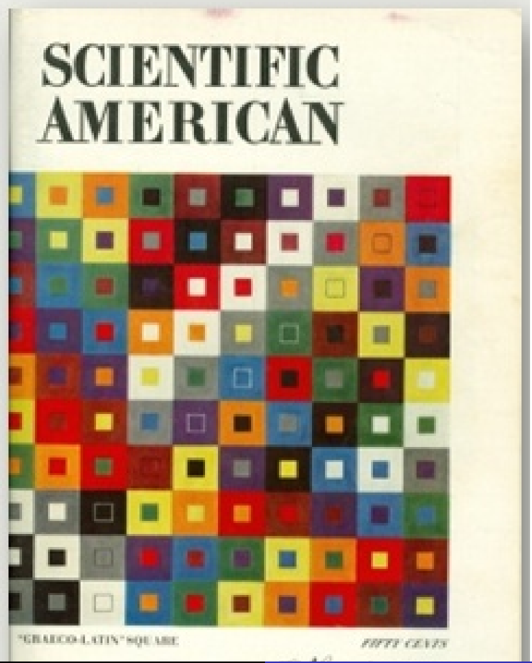
By JOHN A. OSMUNDSEN

Another major mathematical problem—this one 177 years old—has been solved.

New Yorker Hotel yesterday. It was the second such achievement to come out of the meeting, something attending mathematicians called

John G. Thompson, a 26-year-old mathematician from De-Paul University in Chicago. It dealt with so-called "group theory" and had puzzled

Euler (pronounced "oiler") stated it in a memoir in 1782. It became famous as Euler's conjecture. The three mathematicians who finally cracked


SCIENTIFIC
AMERICAN
"GRAECO-LATIN" SQUARE
FIFTY CENTS

Застосовуючи латинські квадрати, за звичай, виходять з того, що ефекти взаємодії між факторами незначні. Тоді результати експерименту можна представити у вигляді лінійної моделі.

Латинський квадрат 3 x 3

A	B			Разом
	b_1	b_2	b_3	
a_1	c_1 y_1	c_2 y_2	c_3 y_3	A_1
a_2	c_2 y_4	c_1 y_5	c_1 y_6	A_2
a_3	c_1 y_7	c_1 y_8	c_2 y_9	A_3
Разом	B_1	B_2	B_3	

Задача Фішера

Припустимо, що необхідно випробувати при мінімальних витратах часу і коштів вплив на зростання пшениці семи сільськогосподарських хімікатів.

Задача Фішера

Припустимо, що необхідно випробувати при мінімальних витратах часу і коштів вплив на зростання пшениці семи сільськогосподарських хімікатів.

Однією з істотних труднощів при випробуваннях такого роду є те, що родючість різних ділянок грунту зазвичай залежить від випадкових факторів.

Задача Фішера

Припустимо, що необхідно випробувати при мінімальних витратах часу і коштів вплив на зростання пшениці семи сільськогосподарських хімікатів.

Однією з істотних труднощів при випробуваннях такого роду ϵ те, що родючість різних ділянок грунту зазвичай залежить від випадкових факторів.

Яким чином можна спланувати експеримент, який дозволить випробувати одночасно всі сім хімікатів і в той же самий час обмежити будь-які сторонні впливи, обумовлені випадковими чинниками?

Відповідь

поділіть пшеничне поле на ділянки, які представлятимуть осередки квадрата зі стороною в сім осередків, потім застосуйте сім "обробок"за моделлю випадково обраного латинського квадрата.

Відповідь

поділіть пшеничне поле на ділянки, які представлятимуть осередки квадрата зі стороною в сім осередків, потім застосуйте сім "обробок"за моделлю випадково обраного латинського квадрата.

Завдяки наявності моделі, простий статистичний аналіз результатів обмежить будь-які помилки, обумовлені випадковими змінами родючості грунту.

А тепер припустимо, що замість одного сорту пшениці необхідно випробувати сім

Чи можна спланувати такий експеримент, який дозволить врахувати ці чотири змінних? (Інші три змінних відображаються родючістю рядів, родючістю колонок і видом обробки.)

А тепер припустимо, що замість одного сорту пшениці необхідно випробувати сім

Чи можна спланувати такий експеримент, який дозволить врахувати ці чотири змінних? (Інші три змінних відображаються родючістю рядів, родючістю колонок і видом обробки.)

Тепер для отримання відповіді використовуєϵться греко-латинський квадрат.

А тепер припустимо, що замість одного сорту пшениці необхідно випробувати сім

Чи можна спланувати такий експеримент, який дозволить врахувати ці чотири змінних? (Інші три змінних відображаються родючістю рядів, родючістю колонок і видом обробки.)

Тепер для отримання відповіді використовується греко-латинський квадрат.

Грецькі літери покажуть, де розмістити сім сортів пшениці, а латинські букви - де застосувати сім різних хімікатів.

А тепер припустимо, що замість одного сорту пшениці необхідно випробувати сім

Чи можна спланувати такий експеримент, який дозволить врахувати ці чотири змінних? (Інші три змінних відображаються родючістю рядів, родючістю колонок і видом обробки.)

Тепер для отримання відповіді використовується греко-латинський квадрат.

Грецькі літери покажуть, де розмістити сім сортів пшениці, а латинські букви - де застосувати сім різних хімікатів.

І в цьому випадку статистичний аналіз результатів не буде представляти складності.

Представимо відхилення, що пов'язане з впливом деякого фактора, від середнього значення досліджуваної величини у вигляді

$$x - \bar{x} = A + e$$

Представимо відхилення, що пов'язане з впливом деякого фактора, від середнього значення досліджуваної величини у вигляді

$$x - \bar{x} = A + e$$

x – конкретне значенні змінної величини,

Представимо відхилення, що пов'язане з впливом деякого фактора, від середнього значення досліджуваної величини у вигляді

$$x - \bar{x} = A + e$$

x – конкретне значенні змінної величини,

$\bar{x}$ – середнϵ значення,

Представимо відхилення, що пов'язане з впливом деякого фактора, від середнього значення досліджуваної величини у вигляді

$$x - \bar{x} = A + e$$

x – конкретне значенні змінної величини,

$\bar{x}$ – середнє значення,

A – частина відхилення змінної, пов'язана з впливом даного фактора,

Представимо відхилення, що пов'язане з впливом деякого фактора, від середнього значення досліджуваної величини у вигляді

$$x - \bar{x} = A + e$$

x – конкретне значенні змінної величини,

$\bar{x}$ – середнє значення,

A – частина відхилення змінної, пов'язана з впливом даного фактора,

e – залишкова частина відхилення, що не пояснюється впливом даного фактора, а задається випадковим впливом.

За ступенем перевищення A над e можна судити про ступінь впливу цього фактора

За ступенем перевищення A над e можна судити про ступінь впливу цього фактора

для цього треба знайти загальну дисперсію

За ступенем перевищення A над e можна судити про ступінь впливу цього фактора

для цього треба знайти загальну дисперсію

$$\sigma_0^2 = \sigma_A^2 + \sigma_e^2$$

За ступенем перевищення A над e можна судити про ступінь впливу цього фактора

для цього треба знайти загальну дисперсію

$$\sigma_0^2 = \sigma_A^2 + \sigma_e^2$$

За аналогією можна розглянути вплив різної кількості факторів

Вплив двох факторів A, B

Вплив двох факторів A, B

$$x - \bar{x} = A + B + AB + e$$

Вплив двох факторів A, B

$$x - \bar{x} = A + B + AB + e$$

$$\sigma_0^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_e^2$$

Вплив двох факторів A, B

$$x - \bar{x} = A + B + AB + e$$

$$\sigma_0^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_e^2$$

Вплив трьох факторів A, B, C

Вплив двох факторів A, B

$$x - \bar{x} = A + B + AB + e$$

$$\sigma_0^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_e^2$$

Вплив трьох факторів A, B, C

$$x - \bar{x} = A + B + C + AB + AC + BC + ABC + e$$

Вплив двох факторів A, B

$$x - \bar{x} = A + B + AB + e$$

$$\sigma_0^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_e^2$$

Вплив трьох факторів A, B, C

$$x - \bar{x} = A + B + C + AB + AC + BC + ABC + e$$

$$\sigma_0^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_{AB}^2 + \sigma_{AC}^2 + \sigma_{BC}^2 + \sigma_{ABC}^2 + \sigma_e^2$$

Вплив двох факторів A, B

$$x - \bar{x} = A + B + AB + e$$

$$\sigma_0^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_e^2$$

Вплив трьох факторів A, B, C

$$x - \bar{x} = A + B + C + AB + AC + BC + ABC + e$$

$$\sigma_0^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_{AB}^2 + \sigma_{AC}^2 + \sigma_{BC}^2 + \sigma_{ABC}^2 + \sigma_e^2$$

Зазвичай кожен із досліджених факторів має не одне, а декілька значень, які називають рівнями або градаціями

Зазвичай кожен із досліджених факторів має не одне, а декілька значень, які називають рівнями або градаціями

Таким чином досліджувані дані можуть бути розбиті на декілька груп, що розрізняються не лише за факторами, а і за їх градаціями

Зазвичай кожен із досліджених факторів має не одне, а декілька значень, які називають рівнями або градаціями

Таким чином досліджувані дані можуть бути розбиті на декілька груп, що розрізняються не лише за факторами, а і за їх градаціями

Дослідження методами дисперсійного аналізу всередені груп, між групами і дисперсії всіх даних вцілому дає можливість встановити вплив даних факторів на мінливість даних

Розглянемо найпростішу схему, коли аналізується вплив лише одного фактора

який може мати декілька рівнів (градацій)

$$1, 2\ldots, i, \ldots, a$$

Розглянемо найпростішу схему, коли аналізується вплив лише одного фактора

який може мати декілька рівнів (градацій)

$$1, 2 \ldots, i, \ldots, a$$

Окремі спостереження розбиваються на групи

$$1, 2 \ldots, j, \ldots, n$$

Розглянемо найпростішу схему, коли аналізується вплив лише одного фактора

який може мати декілька рівнів (градацій)

$$1, 2 \ldots, i, \ldots, a$$

Окремі спостереження розбиваються на групи

$$1, 2 \ldots, j, \ldots, n$$

Розподіл даних за групами представляють у вигляді таблиці

Группы по одному фактору	Отдельные варианты (наблюдения) x_{ij}							Суммы по группам T_i	Средние по группам $\overline{x_i}$
	1	2	3	...	j		n		
1	x_{11}	x_{12}	x_{13}		x_{1j}		x_{1n}	$\Sigma x_1 = T_1$	x_1
2	x_{21}	x_{22}	x_{23}		x_{2j}		x_{2n}	$\Sigma x_2 = T_2$	x_2
:									
:									
:									
i	x_{i1}	x_{i2}	x_{i3}		x_{ij}		x_{in}	$\Sigma x_i = T_i$	$\overline{x_i}$
:									
:									
:									
a	x_{a1}	x_{a2}	x_{a3}		x_{aj}		x_{an}	$\Sigma x_a = T_a$	$\overline{x_a}$
								$\Sigma x_{ij} = T$	$\overline{x}$

Для дослідження мінливості даних зазвичай відокремлюють три напрямки

Для дослідження мінливості даних зазвичай відокремлюють три напрямки

- Загальна зміна всіх даних x_{ij}(незалежно від того, до якої групи вони відносяться) відносно загальної середньої $\bar{x}$

Для дослідження мінливості даних зазвичай відокремлюють три напрямки

- Загальна зміна всіх даних x_{ij}(незалежно від того, до якої групи вони відносяться) відносно загальної середньої $\bar{x}$
- Зміна середніх значень за кожною групою x_i відносно загальної середньої $\bar{x}$

Для дослідження мінливості даних зазвичай відокремлюють три напрямки

- Загальна зміна всіх даних x_{ij}(незалежно від того, до якої групи вони відносяться) відносно загальної середньої $\bar{x}$
- Зміна середніх значень за кожною групою x_i відносно загальної середньої $\bar{x}$
- Зміна всіх даних x_{ij} всередені кожної відносно середнього значення за групою $\bar{x}_i$ (x_i)

Для дослідження мінливості даних зазвичай відокремлюють три напрямки

- Загальна зміна всіх даних x_{ij}(незалежно від того, до якої групи вони відносяться) відносно загальної середньої $\bar{x}$
- Зміна середніх значень за кожною групою x_i відносно загальної середньої $\bar{x}$
- Зміна всіх даних x_{ij} всередені кожної відносно середнього значення за групою $\bar{x}_i$ (x_i)

Для цього треба знайти суму квадратів

Суми квадратів

Суми квадратів

Загальна сума квадратів

$$\sum_{ij}(x_{ij} - \bar{x})^2$$

Суми квадратів

Загальна сума квадратів

$$\sum_{ij}(x_{ij} - \bar{x})^2$$

Сумма квадратів для середніх по групам

$$\sum_i n_i(\bar{x}_i - \bar{x})^2 \qquad n\sum_i(\bar{x}_i - \bar{x})^2$$

Суми квадратів

Загальна сума квадратів

$$\sum_{ij}(x_{ij} - \bar{x})^2$$

Сумма квадратів для середніх по групам

$$\sum_i n_i(\bar{x}_i - \bar{x})^2 \qquad n\sum_i(\bar{x}_i - \bar{x})^2$$

Сумма квадратів відхилень від групових середніх

$$\sum_i\left[\sum_j(x_{ij} - \bar{x}_i)^2\right]$$

Степені свободи

Степені свободи

для загальної дисперсії

$$df = N - 1 \qquad\qquad N = an$$

Степені свободи

для загальної дисперсії

$$df = N - 1 \qquad N = an$$

для дисперсії групових середніх

$$df = a - 1$$

Степені свободи

для загальної дисперсії

$$df = N - 1 \qquad N = an$$

для дисперсії групових середніх

$$df = a - 1$$

для зміни значень всередені групи

$$df = (n-1)a = na - a = N - a$$

Степені свободи

для загальної дисперсії

$$df = N - 1 \qquad N = an$$

для дисперсії групових середніх

$$df = a - 1$$

для зміни значень всередені групи

$$df = (n-1)a = na - a = N - a$$

$$(N-a) + (a-1) = N - a + a - 1 = N - 1$$

Схема однофакторного дисперсійного аналізу

Схема однофакторного дисперсійного аналізу

Сумма квадратов ss	Число степеней свободы df	Средний квадрат ms
$\sum_{ij}(x_{ij}-\overline{x})^2$	$N-1$	$\frac{1}{N-1}\sum_{ij}(x_{ij}-\overline{x})^2$
$\sum_{i} n_i(\overline{x_i}-\overline{x})^2$	$a-1$	$\frac{1}{a-1}\sum_{i} n_i(\overline{x_i}-\overline{x})^2$
$\sum_{i}[\sum_{j}(x_{ij}-\overline{x_i})^2]$	$N-a$	$\frac{1}{N-a}\sum_{i}[\sum_{j} x_{ij}-\overline{x_i})^2]$

Головне в дисперсійному аналізі – це порівняння двох останніх квадратів (другий і третій) з критерієм Фішера

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

Критерій Фішера з рівнем значущості 0,05

Критерій Фішера з рівнем значущості 0,05

df_2 \ df_1	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,94
5	6,61	5,79	5,41	5,19	5,05	4,95	4 88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98

Критерій Фішера з рівнем значущості 0,01

Критерій Фішера з рівнем значущості 0,01

df_2 \ df_1	1	2	3	4	5	6	7	8	9	10
1	4052	4999	5403	5625	5764	5859	5928	5982	6022	6056
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40
3	34,12	30,82	29,46	28,71	28,42	27,91	27,67	27,49	27,35	27,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,25	9,55	8,47	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85

Робочі формули для сум квадратів

Загальна сума квадратів

$$ms = \sigma^2 = \frac{1}{N-1}\left(\sum_{ij} x_{ij}^2 - \frac{T^2}{N}\right)$$

Робочі формули для сум квадратів

Загальна сума квадратів

$$ms = \sigma^2 = \frac{1}{N-1}\left(\sum_{ij} x_{ij}^2 - \frac{T^2}{N}\right)$$

Сумма квадратів для середніх по групам

$$ms = \sigma^2 = \frac{1}{a-1}\left(\sum_{i} \frac{T_i^2}{n_i} - \frac{T^2}{N}\right)$$

Робочі формули для сум квадратів

Загальна сума квадратів

$$ms = \sigma^2 = \frac{1}{N-1}\left(\sum_{ij} x_{ij}^2 - \frac{T^2}{N}\right)$$

Сумма квадратів для середніх по групам

$$ms = \sigma^2 = \frac{1}{a-1}\left(\sum_{i} \frac{T_i^2}{n_i} - \frac{T^2}{N}\right)$$

Сумма квадратів випадкових відхилень

$$ms = \sigma^2 = \frac{1}{N-a}\left(\sum_{ij} x_{ij}^2 - \sum_{i} \frac{T_i^2}{n_i}\right)$$

Приклад дисперсійного аналізу для одно факторної схеми та однакової кількості даних у кожній групі

Приклад дисперсійного аналізу для одно факторної схеми та однакової кількості даних у кожній групі

Часы суток	Определение x_{ij} 1	2	3	4	T_i	n_i	$\overline{x}_i$	T_i^2
15	1,41	0,95	1,00	0,93	4,29	4	1,07	18,4041
18	1,17	1,10	0 84	1,01	4,12	4	1,03	16,9744
21	1,38	1,38	0,91	1,36	5,03	4	1,26	25,3009
24	0,62	0,48	0,43	0,62	2,15	4	0,54	4,6225
6	0,74	0,41	0,41	0,43	1,96	4	0,50	3,9601
9	0,76	0,59	0,74	0,46	2,55	4	0,64	6 5025
12	0,64	1,02	1,04	0,98	3,68	4	0,92	13,5422
					$T = 23{,}81$ $T^2 = 566{,}9161$	$N=28$		$\Sigma T_i^2=89{,}3069$

Окремо обчислимо

Окремо обчислимо

$$\Sigma x_{ij}^2 = 22{,}7316$$

Окремо обчислимо

$$\Sigma x_{ij}^2 = 22,7316$$

$$\sum_{ij} x_{ij}^2 - \frac{T^2}{N} = 22,7316 - \frac{566,9161}{28} =$$

Окремо обчислимо

$$\Sigma x_{ij}^{2} = 22{,}7316$$

$$\sum_{ij} x_{ij}^{2} - \frac{T^{2}}{N} = 22{,}7316 - \frac{566{,}9161}{28} =$$

$$= 22{,}7316 - 20{,}2470 = 2{,}4846.$$

Окремо обчислимо

$$\Sigma x_{ij}^2 = 22,7316$$

$$\sum_{ij} x_{ij}^2 - \frac{T^2}{N} = 22,7316 - \frac{566,9161}{28} =$$

$$= 22,7316 - 20,2470 = 2,4846.$$

$$\sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$\sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N} = \frac{1}{n} \Sigma T_i^2 - \frac{T^2}{N} =$$

$$\sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N} = \frac{1}{n} \sum T_i^2 - \frac{T^2}{N} =$$

$$= \frac{1}{4} \cdot 89{,}3069 - 20{,}2470 =$$

$$\sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N} = \frac{1}{n}\sum T_i^2 - \frac{T^2}{N} =$$

$$= \frac{1}{4} \cdot 89{,}3069 - 20{,}2470 =$$

$$= 22{,}3267 - 20{,}2470 = 2{,}0797.$$

$$\sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N} = \frac{1}{n}\Sigma T_i^2 - \frac{T^2}{N} =$$

$$= \frac{1}{4} \cdot 89{,}3069 - 20{,}2470 =$$

$$= 22{,}3267 - 20{,}2470 = 2{,}0797.$$

$$\sum_{ij} x_{ij}^2 - \sum \frac{T_i^2}{n_i} = 22{,}7316 - 22{,}3267 = 0{,}4049.$$

$$df = N - 1 = 28 - 1 = 27;$$
$$df = a - 1 = 7 - 1 = 6;$$
$$df = N - a = 28 - 7 = 21.$$

$$df = N - 1 = 28 - 1 = 27;$$
$$df = a - 1 = 7 - 1 = 6;$$
$$df = N - a = 28 - 7 = 21.$$

Источник варьирования	Сумма квадратов ss	Число степеней свободы df	Средний квадрат ms	F фактическое	F табличное	
					при $P=0{,}05$	при $P=0{,}01$
Общее	2,4846	27	—			
Фактор A (время суток)	2,0797	6	0,3466	$\frac{0{,}3466}{0{,}0193} = 18{,}0$	2,57	3,81
Случайные отклонения	0,4049	21	0,0193			