ДИСПЕРСІЙНИЙ АНАЛІЗ

Основною метою дисперсійного аналізу, фундаментальна концепція якого була запропонована Фішером у 1920 р., є дослідження значущості відмінності між середніми декількох груп даних або змінних. Якщо порівнюються середні двох груп, дисперсійний аналіз дасть той же результат, що і звичайний ґ-критерій для незалежних або залежних вибірок. Проте використання дисперсійного аналізу має переваги особливо для малих вибірок.

У дисперсійному аналізі перевірка статистичної значущості відмінності між середніми декількох груп здійснюється на основі вибіркових дисперсій. Ця перевірка проводиться за допомогою розбиття загальної дисперсії (варіації) на частини, одна з яких обумовлена випадковою помилкою (тобто внутрішньогруповою мінливістю), а друга пов'язана з відмінністю середніх значень. Якщо ця відмінність значуща, нульова гіпотеза щодо існування відмінності між середніми значеннями відкидається на певному рівні значущості.

Дисперсійний однофакторний аналіз

Дисперсійний однофакторний аналіз використовується у дослідженнях зміни результативної ознаки під впливом зміни умов або градацій фактора. Суть математичних перетворень дисперсійного методу полягає в тому, щоб зіставити дисперсії за факторами із дисперсією усіх значень, отриманих в експерименті. Однофакторний аналіз вимагає не менше трьох градацій фактора і не менше двох випробовувань у кожній градації. При проведенні дисперсійного аналізу необхідно перевірити нормальність розподілу досліджуваної випадкової величини і відсутність відмінності дисперсій сукупностей. Це можна виконати методами перевірки статистичних гіпотез.

Припустимо, що аналізується вплив фактора A на к рівнях A_1, ^4$_2$, $A_к$. Наприклад, в експерименті це можна реалізувати, якщо задіяти к вибірок з різними градаціями умов. На кожному рівні *Лі* (для кожної вибірки) проведено *п* спостережень$x_{/1}$, $x_{/2}$, $x_{іп}$ (див. табл. 6.1).

Таблиця 6.1

Номери спостережень	Рівні фактора A		
		A2	*$A_к$*
1	^{x}и	^{x}21 ооо	*хи*
2		^{x}22	*хк2*
I	^{x}1і	^{x}2I о	^{x}I
п	^{x}1п	^{x}2п	^{x}кп
	X1	*X2 о..*	*Xк*

Розглянемо оцінки різних дисперсій.

Дисперсія я^2 для рівня Ai (для певної вибірки) може бути записана як

$$s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_{ij}\right)^2\right].$$

Дисперсія я0 , що характеризує варіативність поза впливу фактора *А*

$$s_0^2 = \frac{1}{k}\sum_{i=1}^{k} s_i^2 = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2 = \frac{1}{k(n-1)}\left[\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{n}\sum_{i=1}^{k}\left(\sum_{j=1}^{n} x_{ij}\right)^2\right].$$

Загальна дисперсія я всіх *пк* спостережень дорівнює

$$s^2 = \frac{1}{kn-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{\bar{x}})^2, \text{ де } \bar{\bar{x}} = \frac{1}{k}\sum_{i=1}^{k}\bar{x}_i \text{ і } \bar{x}_i = \frac{1}{k}\sum_{j=1}^{n} x_{ij}.$$

Отже,

$$s_2 = \frac{1}{kn-1}\left[\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{kn}\left(\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}\right)^2\right].$$

Дисперсія я $^2{}_A$, що характеризує зміну середніх х/ під впливом фактора А:

1 к _ =

к ~ 1 ¡=1

Перевірка впливу фактора А на зміну середніх може бути зведена до порівняння дисперсій я$^2{}_A$ і я^2 . Вплив фактора А вважатиметься значущим на ріні

а, якщо є значущим відношення 5 1я^2, тобто якщо

- 5$^2{}_Л$ і^ > ^а[к -1; к(п -1)], де *к* -1; к(п -1) - ступені вільності ^-розподілу, 5 і *я*^{7}*]* - ^-критерій Фішера. Приклад 6.1. Довести припущення про те, що фактор швидкості пред'явлення слів впливає на показники їх відтворення (дані у таблиці рис. 8.1). *Послідовність рішення*:
  - о *Формулювання гіпотез.*

Н0: фактор швидкості не є більш вираженим, ніж випадковим; Н1: фактор швидкості є більш вираженим, ніж випадковим.

- о *Перевірка припущень:* досліджуваний параметр має *нормальний* розподіл; вибірки *незв'язані однакових* обсягів; виміри за шкалою відношень.
- о *Визначення емпіричного критерію Геми* базується на зіставленні квадратів сум по стовпцях із сумою квадратів усіх емпіричних значень. Кожний стовпець представляє вибірку і відповідає певній градації фактора швидкості.
- о *Введені позначення*:

п = 6 - кількість спостережень (рядків);

$\kappa = 3$ - кількість факторів (стовпчиків);

$n\kappa$ = 6-3 = 18 - загальна кількість індивідуальних значень;

7 - індекс рядків змінюється від 1 до *n* (7 =1, 2, ..., п);

i - індекс стовпчиків змінюється від 1 до *к (i* =1, 2, ..., к).

- о *Математичні розрахунки* (див. рис 6.1 і 6.2):
- \- розрахувати суми в комірках В13:В15 за формулами

і=1 *7*=1 *n* *м* *кп* ^ і=1)

а саме

Є1 = $6^2 + 7^2 + 6^2 + 5^2$ +_+5^2 +5^2 = 432; й2 = --($34^2 + 29^2 + 23^2$) = 421;

6

й3 ^^$(34 + 29 + 23)^2$ = 410,89; 3 о 6

\- розрахувати емпіричний критерій ¥гмп в комірці В16 за формулою

$$F_{емп} = \frac{s_A^2}{s_0^2} = \frac{k(n-1)}{k-1}\frac{Q_2 - Q_3}{Q_1 - Q_2}, \text{ тобто } F_{емп} = \frac{3(6-1)}{3-1}\frac{421-410{,}89}{432-421} \approx 6{,}89.$$

	A	B	C	D
1	j \ i	Швидкість пред'явлення		
2		Низька	Середня	Висока
3	1	6	5	4
4	2	7	6	4
5	3	6	5	4
6	4	5	4	3
7	5	6	4	5
8	6	4	5	3
9	Суми:	34	29	23
10	Середні:	5,67	4,83	3,83
11	n =	6		
12	k =	3		
13	$Q1$ =	432,00		
14	$Q2$ =	421,00		
15	$Q3$ =	410,89		
16	$F_{емп}$ =	6,89		
17	$F_{0,05}$ =	3,68		
18	$F_{0,01}$ =	6,36		
19	$p_{емп}$ =	0,0075		

	A	B	C	D
1	j \ i	Швидкість пред'явлення		
2		Низька	Середня	Висока
3	1	6	5	4
4	2	7	6	4
5	3	6	5	4
6	4	5	4	3
7	5	6	4	5
8	6	4	5	3
9	Суми:	=СУММ(B3:B8)	=СУММ(C3:C8)	=СУММ(D3:D8)
10	Середні:	=СРЗНАЧ(B3:B8)	=СРЗНАЧ(C3:C8)	=СРЗНАЧ(D3:D8)
11	n =	=СЧЁТ(B3:B8)		
12	k =	=СЧЁТ(B9:D9)		
13	$Q1$ =	=СУММКВ(B3:D		
14	$Q2$ =	=СУММКВ(B9:D9)/B11		
15	$Q3$ =	=СУММ(B9:D9)^2/B11/B12		
16	$F_{емп}$ =	=(B14-B15)*B12*(B11-1)/(B13-B14)/(B12-1)		
17	$F_{0,05}$ =	=FРАСПОБР(0,05;B12-1;B12*(B11-1))		
18	$F_{0,01}$ =	=FРАСПОБР(0,01;B12-1;B12*(B11-1))		
19	$p_{емп}$ =	=FРАСП(B16;B12-1;B12*(B11-1))		

Рис. 6.1. Результати Рис. 6.2. Розрахункові формули

дисперсійного аналізу однофакторного дисперсійного аналізу

о Критичне значення ^кр можна отримати за допомогою функції

=РРАСПОБР() для рівня значущості для a = 0,05 (0,01) і числа ступенів вільності *к* -1 = 3-1=2 і к(п -1) = 3(6-1)=15. $Г_{0i05}$ ~ 3,68 і $Г_{0i01}$ ~ 6,36.

- о *Прийняття рішення.* Оскільки ¥$_{гмп}$ >$P_{0?01}$ (6,89 > 6,36), нульова гіпотеза H_0 відхиляється на рівні значущості 0,01.
- о *Формулювання висновків.* Відмінності в обсязі відтворення слів (фактор швидкості) є більш вираженими, ніж випадковим. Цю залежність можна представити графічно на рис. 6.3.

Рис. 6.3. Залежність середнього обсягу відтворених слів від швидкості пред'явлення

Розрахунки однофакторної моделі можна провести за допомогою пакета "Аналіз даних" розділ "Однофакторний дисперсійний аналіз" (рис. 6.4).

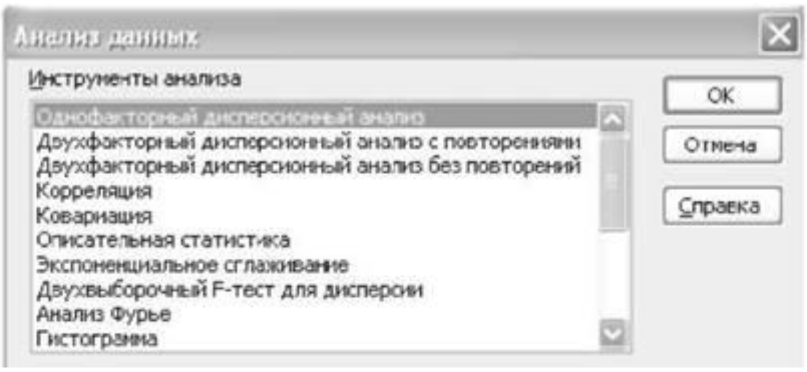

Рис. 6.4. Меню пакета "Аналіз даних" Після введення відповідних параметрів (рис. 6.5) можна отримати результати однофакторного дисперсійного аналізу (рис. 6.6).

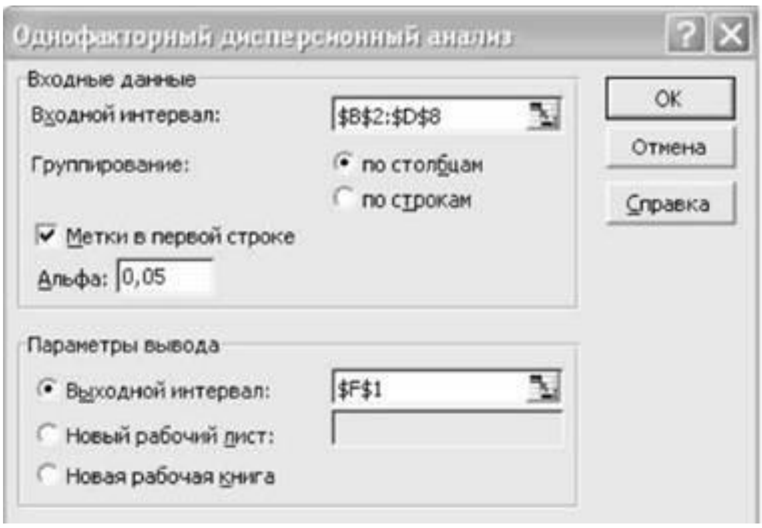

Рис. 6.5. Діалогове вікно

	F	G	H	I	J	K	L
1	Однофакторный дисперсионный анализ						
2							
3	ИТОГИ						
4	*Группы*	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
5	Низька	6	34	5,67	1,07		
6	Середня	6	29	4,83	0,57		
7	Висока	6	23	3,83	0,57		
8							
9							
10	Дисперсионный анализ						
11	*Источник вариации*	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
12	Между группами	10,11	2	5,06	6,89	0,0075	3,68
13	Внутри групп	11	15	0,73			
14							
15	Итого	21,11	17				

Рис. 6.6. Результати однофакторного дисперсійного аналізу (a = 0,05)