

Методичні рекомендації

Однофакторний дисперсійний аналіз для незалежних вибірок

Гіпотези:

H_0 : Відмінності між градаціями фактора (різними умовами) не більш вираженими, ніж випадкові відмінності всередині кожної групи.

H_1 : Відмінність між градаціями фактора (різними умовами) є більш вираженими, ніж випадкові відмінності всередині кожної групи.

Алгоритм:

1. Сформулювати гіпотезу.
2. Підготувати таблицю для обчислень:

i	Фактор 1		Фактор 2		...	Фактор k		$\Sigma(x_1+x_2+\dots+x_k)$
	x_1	x_1^2	x_2	x_2^2	...	x_k	x_k^2	
1	~~
2	~~
3	~~
4	~~
5	~~
...	~~
Σ	Σx_1	$\Sigma(x_1)^2$	Σx_2	$\Sigma(x_2)^2$...	Σx_k	$\Sigma(x_k)^2$	Σx_{total}

3. Обчислити суми квадратів.

$$\text{Загальну} - SS_{total} = \Sigma(x_1)^2 + \Sigma(x_2)^2 + \dots + \Sigma(x_k)^2 - \frac{(\Sigma x_{total})^2}{N}$$

$$\text{Міжгрупову} - SS_{bg} = \frac{\Sigma(x_1)^2}{n_1} + \frac{\Sigma(x_2)^2}{n_2} + \dots + \frac{\Sigma(x_k)^2}{n_k} - \frac{(\Sigma x_{total})^2}{N}$$

$$\text{Внутрішньогрупову} - SS_{wg} = SS_{total} - SS_{bg}$$

1. Визначити кількість ступенів свободи df : $df_{total} = N - 1$; $df_{bg} = k - 1$;
 $df_{wg} = df_{total} - df_{bg} = N - k$.

2. Обчислити середні квадрати:

$$\text{Внутрішньогруповий} - MS_{wg} = \frac{SS_{wg}}{df_{wg}};$$

$$\text{Міжгруповий} - MS_{bg} = \frac{SS_{bg}}{df_{bg}}$$

3. Обчислити $F_{емп}$:

$$F_{\text{емп}} = \frac{MS_{bg}}{MS_{wg}}$$

4. За таблицями критичних значень F-критерію Фішера для $df_{bg}=k-1$ (чисельник) та df_{wg} (знаменник) знайти $F_{кр1}$ ($p=0,05$) и $F_{кр2}$ ($p=0,01$).
5. Побудувати вісь значущості та визначити зону, в яку потрапило $F_{\text{емп}}$.
6. Сформулювати рішення щодо гіпотези, що приймається.
7. Побудувати графік середніх.
8. Обчислити коефіцієнт детермінації:

$$R^2 = \frac{SS_{bg}}{SS_{total}}$$

Однофакторний дисперсійний аналіз для залежних вибірок

Гіпотези:

H_{0A}: Внутрішньогрупові відмінності між градаціями фактора не більш виражені, ніж відмінності, зумовлені випадковими причинами;

H_{1A}: Внутрішньогрупові різницю між градаціями чинника більш виражені, ніж відмінності, обумовлені випадковими причинами.

H_{0B}: Міжгрупові відмінності між градаціями фактора є не більш вираженими, ніж відмінності, зумовлені випадковими причинами;

H_{1B}: Міжгрупові різницю між градаціями фактора є більш вираженими, ніж відмінності, обумовлені випадковими причинами.

Алгоритм:

1. Сформулювати гіпотези.
2. Підготувати таблицю для обчислень:

i	Фактор 1		Фактор 2		...	Фактор k		$\Sigma(x_1+x_2+\dots+x_k)$
	x_1	x_1^2	x_2	x_2^2		x_k	x_k^2	
1	~~
2	~~
3	~~
4	~~
5	~~
...	~~
Σ	Σx_1	$\Sigma(x_1)^2$	Σx_2	$\Sigma(x_2)^2$...	Σx_k	$\Sigma(x_k)^2$	Σx_{total}

3. Обчислити суми квадратів.

$$\text{Загальну} - SS_{total} = \Sigma(x_1)^2 + \Sigma(x_2)^2 + \dots + \Sigma(x_k)^2 - \frac{(\Sigma x_{total})^2}{N}$$

$$\text{Міжгрупову} - SS_{bg} = \frac{\Sigma(x_1)^2 + \Sigma(x_2)^2 + \dots + \Sigma(x_k)^2}{n} - \frac{(\Sigma x_{total})^2}{N}$$

$$\text{Внутрішньогрупову} - SS_{wg} = \frac{\Sigma(\Sigma x_i)^2}{k} - \frac{(\Sigma x_{total})^2}{N}$$

$$\text{Випадкову} - SS_{casual} = SS_{total} - SS_{bg} - SS_{wg}$$

4. Визначити кількість ступенів свободи df : $df_{total} = N - 1$; $df_{bg} = k - 1$;
 $df_{wg} = n - 1$, $df_{casual} = df_{total} - df_{bg} - df_{wg}$.

5. Обчислити середні квадрати:

$$\text{Внутрішньогруповий} - MS_{wg} = \frac{SS_{wg}}{df_{wg}};$$

$$\text{Міжгруповий} - MS_{wbg} = \frac{SS_{bg}}{df_{bg}};$$

$$\text{Випадковий} - MS_{casual} = \frac{SS_{casual}}{df_{casual}}$$

6. Обчислити $F_{емп}$:

$$\text{Міжгрупове} - F_{bg \text{ емп}} = \frac{MS_{bg}}{MS_{casual}};$$

$$\text{Внутрішньогрупове} - F_{wg \text{ емп}} = \frac{MS_{wg}}{MS_{casual}}.$$

7. За таблицями критичних значень F-критерію Фішера для df_{bg} (числитель) и df_{casual} (знаменатель), df_{wg} (числитель) и df_{casual} (знаменатель) найти $F_{кр1}$ и $F_{кр2}$.

8. Побудувати осі значущості та визначити зони, в які потрапили $F_{bg \text{ емп}}$ та $F_{wg \text{ емп}}$.

9. Сформулювати рішення про прийняті гіпотези.

10. Побудувати графік середніх.

11. Обчислити коефіцієнти детермінації:

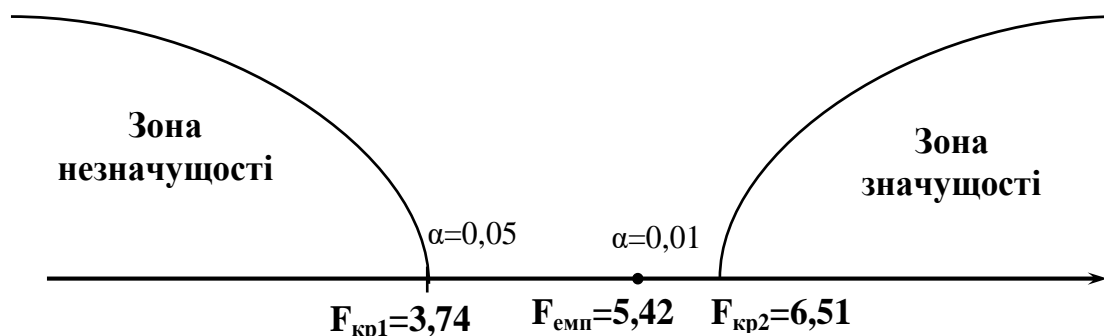
$$R_{bg}^2 = \frac{SS_{bg}}{SS_{total}}, R_{wg}^2 = \frac{SS_{wg}}{SS_{total}}, R_{casual}^2 = \frac{SS_{casual}}{SS_{total}}.$$

3. Обчислити суми квадратів (SS_{total} , SS_{bg} та SS_{wg}), число ступенів свободи, середні квадрати, $F_{емп}$ та коефіцієнт детермінації R^2 .

	A	B	C		A	B	
7	5	7	49	7	5	7	=B7^
8	6	5	25	8	6	5	=B8^
9	Σ	30	160	9	Σ	=СУММ(B3:B8)	=СУ
10	n_i =	6		10	n_i =	=СЧЁТ(B3:B8)	
11	k =	3		11	k =	3	
12	N =	17		12	N =	=СУММ(B10:F10)	
13	SS_{total} =	63,88		13	SS_{total} =	=C9+E9+G9-H9^2/B12	
14	SS_{bg} =	27,88		14	SS_{bg} =	=B9^2/B10+D9^2/D10+F9^2/F10-H9^2/B12	
15	SS_{wg} =	36,00		15	SS_{wg} =	=B13-B14	
16	df_{bg} =	2		16	df_{bg} =	=B11-1	
17	df_{wg} =	14		17	df_{wg} =	=B12-B11	
18	MS_{bg} =	13,94		18	MS_{bg} =	=B14/B16	
19	MS_{wg} =	2,57		19	MS_{wg} =	=B15/B17	
20	$F_{емп}$ =	5,42		20	$F_{емп}$ =	=B18/B19	
21	$F_{кр1}$ =	3,74		21	$F_{кр1}$ =	=F.ОБР.ПХ(0,05;B16;B17)	
22	$F_{кр2}$ =	6,51		22	$F_{кр2}$ =	=F.ОБР.ПХ(0,01;B16;B17)	
23	R^2 =	0,44		23	R^2 =	=B14/B13	
24				24			

4. Обчислені критичні значення $F_{кр1} = 3,74$ і $F_{кр2} = 6,51$.

5. Побудуємо вісь значущості:



6. $F_{емп}$ знаходиться в зоні невизначеності, тому порівняємо $F_{емп}$ з $F_{кр1}$. Т.к. $F_{емп} > F_{кр1}$, то на рівні $\alpha = 0,05$ гіпотеза H_0 відхиляється і приймається гіпотеза H_1 . Наявний вплив фактора стажу роботи на ступінь задоволеності роботою: $R^2 = 0,44$. Це означає, що 44% варіабельності даних обумовлено факторним ефектом (збільшення стажу роботи).

7. Скористаємося інструментом **Однофакторний дисперсійний аналіз** для отримання $F_{емп}$:

	A	B	C	D
1		Стаж роботи		
2	i	0-5 років	6-10 років	11-15 років
3	1	5	8	6
4	2	3	9	5
5	3	6	10	9
6	4	4	9	7
7	5	7	5	8
8	6	5	7	
9				

Однофакторный дисперсионный анализ ? X

Входные данные

Входной интервал:

Группирование: по столбцам по строкам

Метки в первой строке

Альфа:

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

	F	G	H	I	J	K	L	M
1	Однофакторный дисперсионный анализ							
2								
3	Підсумки							
4	<i>Групи</i>	<i>Кількість</i>	<i>Сума</i>	<i>Середнє</i>	<i>Дисперсія</i>			
5	0-5 років	6	30	5	2			
6	6-10 років	6	48	8	3,2			
7	11-15 років	5	35	7	2,5			
8								
9								
10	Дисперсійний аналіз							
11	<i>Джерело варіації</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>α</i>	<i>F_{кр}</i>	
12	Між групами	27,88	2	13,94	5,42	0,02	3,74	
13	Всередині груп	36	14	2,57				
14								
15	Всього	63,882	16					
16								
17								

8. На основі середніх, отриманих в ході застосування однофакторного дисперсійного аналізу (діапазон I5:I7), побудований графік середніх (додані планки похибок):

