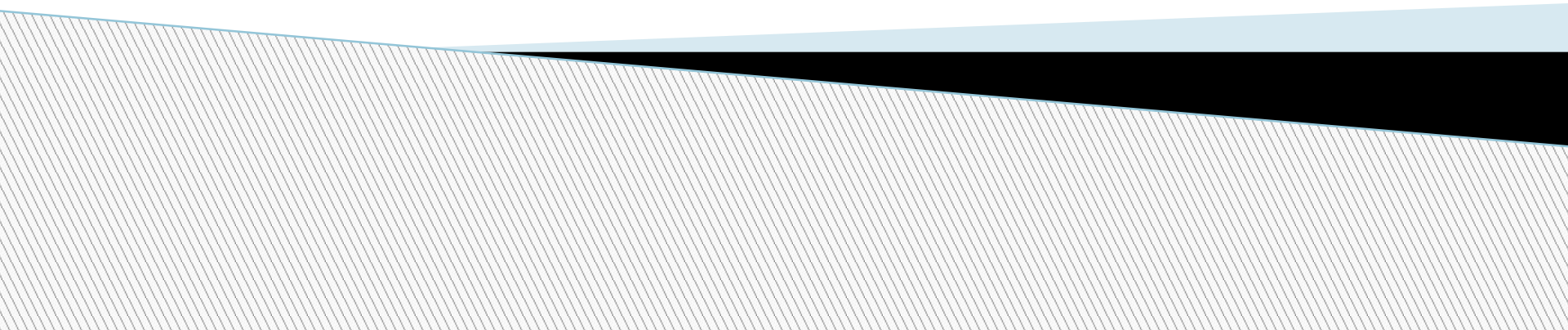


**ОП.02 ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА**

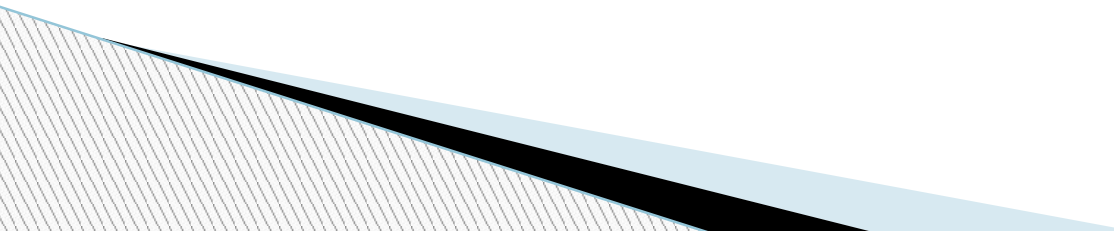
Тема «Двухфакторный дисперсионный анализ»



Задача дисперсионного анализа

- Допустим, что нас интересует зависимость результативного признака Y , который является случайной величиной, от двух факторов A и B .

Основные понятия

- A_1, A_2, \dots, A_{v_A} - уровни фактора A
 - v_A - всего уровней фактора A
 - B_1, B_2, \dots, B_{v_B} - уровни фактора B
 - v_B - всего уровней фактора B
 - $n = v_A v_B$ - общее число наблюдений
- 

Результаты наблюдений и результаты их предварительной обработки расположим в таблице 1

Таблица 1.

		Уровни фактора B		B_1	B_2	B_{v_B}	Среднее групповое (по строке)
Уровни фактора A	i	j	1	2	...	v_B		
	A_1	1	Y_{11}	Y_{12}	...	Y_{1v_B}	$\bar{Y}_{1\cdot}$ $\bar{Y}_{2\cdot}$	
A_2	2	Y_{21}	Y_{22}	...	Y_{2v_B}			
.....		
A_{v_A}	v_A	$Y_{v_A 1}$	$Y_{v_A 2}$...	$Y_{v_A v_B}$	$\bar{Y}_{v_A \cdot}$		
Среднее групповое (по столбцу)			$\bar{Y}_{\cdot 1}$	$\bar{Y}_{\cdot 2}$		$\bar{Y}_{\cdot v_B}$		

- Y_{ij} - результат наблюдений, зафиксированный на i -м уровне фактора A и j -м уровне фактора B .

Гипотеза:

- Необходимо проверить зависимость результата от выбранного фактора.

a_i – математическое ожидание при уровне A_i ,

b_j – математическое ожидание при уровне B_j

- Если при изменении уровня фактора A сохраняется равенство $a_1 = a_2 = \dots = a_{VA}$, то Y не зависит от фактора A
- Если при изменении уровня фактора B сохраняется равенство $b_1 = b_2 = \dots = b_{VB}$, то Y не зависит от фактора B .

Этапы проверки гипотезы

1. Предварительная обработка результатов наблюдений:

1.1. Формула группового среднего (по строке, фактор А):

$$\bar{Y}_{i\cdot} = \frac{Y_{i1} + Y_{i2} + \dots + Y_{iv_B}}{v_B}, \text{ где } i = 1 \dots v_A$$

1.2. Формула группового среднего (по столбцу, фактор В):

$$\bar{Y}_{\cdot j} = \frac{Y_{1j} + Y_{2j} + \dots + Y_{v_A j}}{v_A}, \text{ где } j = 1 \dots v_B$$

1.3. Формула общего среднего: $\bar{Y} = \frac{\sum_{i=1}^{v_A} \sum_{j=1}^{v_B} Y_{ij}}{n}$

2. Разложение дисперсии:

2.1. Формулы общей вариации: $S_Y^2 = \sum_{i=1}^{v_A} \sum_{j=1}^{v_B} (Y_{ij} - \bar{Y})^2$; $\overline{\sigma_Y^2} = \frac{S_Y^2}{n}$

2.2. Формулы A факторной вариации (по строке): $S_A^2 = \sum_{i=1}^{v_A} (\bar{Y}_{i\cdot} - \bar{Y})^2 \cdot v_B$; $\overline{\sigma_A^2} = \frac{S_A^2}{n}$

2.3. Формулы B факторной вариации (по столбцу): $S_B^2 = \sum_{j=1}^{v_B} (\bar{Y}_{\cdot j} - \bar{Y})^2 \cdot v_A$; $\overline{\sigma_B^2} = \frac{S_B^2}{n}$

2.4. Формулы остаточной вариации: $S_o^2 = \sum_{i=1}^{v_A} \sum_{j=1}^{v_B} (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y})^2$; $\overline{\sigma_o^2} = \frac{S_o^2}{n}$

3. Таблица двухфакторного дисперсионного анализа:

Источник	Показатель вариации	Число степеней свободы	Несмещенная оценка дисперсии
Фактор А	$\overline{\sigma_A^2}$	$v_A - 1$	$s_A^2 = \frac{S_A^2}{v_A - 1}$
Фактор В	$\overline{\sigma_B^2}$	$v_B - 1$	$s_B^2 = \frac{S_B^2}{v_B - 1}$
Остаточный фактор	$\overline{\sigma_o^2}$	$(v_A - 1) \cdot (v_B - 1)$	$s_o^2 = \frac{S_o^2}{(v_A - 1) \cdot (v_B - 1)}$
Общая вариация	$\overline{\sigma_Y^2}$	$n - 1$	$s_Y^2 = \frac{S_Y^2}{n - 1}$

$$\overline{\sigma_Y^2} = \overline{\sigma_A^2} + \overline{\sigma_B^2} + \overline{\sigma_o^2}$$

4. Проверка основной гипотезы

4.1. Фактора A

- Формула вычисления критерия: $F^A = \frac{s_A^2}{s_0^2}$
- Схема нахождения критической области:

$$\left. \begin{array}{l} \alpha \\ k_1 = v_A - 1, \\ k_2 = (v_A - 1) \cdot (v_B - 1) \end{array} \right\} \text{П6 } f_Y = X_{\text{пр},\alpha}^{\text{кр}} \rightarrow \omega = (X_{\text{пр},\alpha}^{\text{кр}}; +\infty)$$

- Принятие решения: $\varphi \in \omega \rightarrow H_1$
 $\varphi \notin \omega \rightarrow H_0$

4.2. Фактора B

- Формула вычисления критерия: $F^B = \frac{s_B^2}{s_0^2}$
- Схема нахождения критической области:

$$\left. \begin{array}{l} \alpha \\ k_1 = v_B - 1, \\ k_2 = (v_A - 1) \cdot (v_B - 1) \end{array} \right\} \text{П6 } f_Y = X_{\text{пр},\alpha}^{\text{кр}} \rightarrow \omega = (X_{\text{пр},\alpha}^{\text{кр}}; +\infty)$$

- Принятие решения: $\varphi \in \omega \rightarrow H_1$
 $\varphi \notin \omega \rightarrow H_0$

5. Степень влияния фактора на результативный признак (формула выборочного коэффициента детерминации):

$$\overline{\rho^2}_A = \frac{\overline{\sigma_A^2}}{\sigma_Y^2} \leq 1;$$

$$\overline{\rho^2}_B = \frac{\overline{\sigma_B^2}}{\sigma_Y^2} \leq 1$$

Задача:

- Выяснить при уровне значимости 0,05, влияют ли качество пряжи, измеряемое величиной разрывной нагрузки, тип машины и вид сырья, из которого пряжа производится. Необходимые данные приведены в таблице 2.

Таблица 2

A \ B	B₁	B₂	<i>Среднее групповое по строке</i>		
A₁	$Y_{11}=10$	$Y_{12}=50$			
A₂	$Y_{21}=20$	$Y_{22}=60$			
A₃	$Y_{31}=30$	$Y_{32}=100$			
<i>Среднее групповое по столбцу</i>					

ЭТАПЫ

1. Предварительная обработка результатов наблюдений:

1.1. Групповое среднее (по строке, фактор А):

$$\bar{Y}_{1.} = \frac{10+50}{2} = 30, \quad \bar{Y}_{2.} = \frac{20+60}{2} = 40, \quad \bar{Y}_{3.} = \frac{30+100}{2} = 65$$

1.2. Групповое среднее (по столбцу, фактор В):

$$\bar{Y}_{.1} = \frac{10+20+30}{3} = 20, \quad \bar{Y}_{.2} = \frac{50+60+100}{3} = 70$$

1.3. Формула общего среднего:

$$\bar{Y} = \frac{10+50+20+60+30+100}{6} = \frac{20*3+70*3}{6} = \frac{30*2+40*2+65*2}{6} = 45$$

A \ B	B1	B2	<i>Среднее групповое по строке</i>
A1	$Y_{11}=10$	$Y_{12}=50$	$\bar{Y}_{1.} = 30$
A2	$Y_{21}=20$	$Y_{22}=60$	$\bar{Y}_{2.} = 40$
A3	$Y_{31}=30$	$Y_{32}=100$	$\bar{Y}_{3.} = 65$
<i>Среднее групповое по столбцу</i>	$\bar{Y}_{.1} = 20$	$\bar{Y}_{.2} = 70$	$\bar{Y} = 45$

2. Разложение дисперсии:

2.1. Общая вариация:

$$S_Y^2 = (10 - 45)^2 + (50 - 45)^2 + (20 - 45)^2 + (60 - 45)^2 + (30 - 45)^2 + (100 - 45)^2 = 5350$$

$$\overline{\sigma_Y^2} = \frac{5350}{6} = 891,67$$

2.2. Вариация, обусловленная фактором А:

$$S_A^2 = (30 - 45)^2 * 2 + (40 - 45)^2 * 2 + (65 - 45)^2 * 2 = 1300,$$

$$\overline{\sigma_A^2} = \frac{1300}{6} = 216,67$$

2.3. Вариация, обусловленная фактором В:

$$S_B^2 = (20 - 45)^2 * 3 + (70 - 45)^2 * 3 = 3750,$$

$$\overline{\sigma_B^2} = \frac{3750}{6} = 625$$

2.4. Остаточная вариация:

$$S_o^2 = (10 - 30 - 20 + 45)^2 + (50 - 30 - 70 + 45)^2 + (20 - 40 - 20 + 45)^2 + \\ + (60 - 40 - 70 + 45)^2 + (30 - 65 - 20 + 45)^2 + (100 - 65 - 70 + 45)^2 = 300$$

$$\overline{\sigma_o^2} = \frac{300}{6} = 50$$

3. Таблица двухфакторного дисперсионного анализа: $\bar{\sigma}_Y^2 = \bar{\sigma}_A^2 + \bar{\sigma}_B^2 + \bar{\sigma}_o^2$

Источник	Показатель вариации	Число степеней свободы	Несмещенная оценка дисперсии
Фактор А (тип машины)	216,67	2	$s_A^2 = \frac{1300}{2} = 650$
Фактор В (вид сырья)	625	1	$s_B^2 = \frac{3750}{1} = 3750$
Остаточный фактор	50	2	$s_o^2 = \frac{300}{2} = 150$
Общая вариация	891,67	5	$s_Y^2 = \frac{5350}{5} = 1070$

4. Проверка гипотез

4.1. Фактора A

- Формула вычисления критерия: $F^A = \frac{s_A^2}{s_0^2} = \frac{650}{150} = 4,3$

- Схема нахождения критической области:

$$\left. \begin{array}{l} \alpha = 0,05 \\ k_1 = v_A - 1 = 2, \\ k_2 = (v_A - 1) \cdot (v_B - 1) = 2 \end{array} \right\} \text{П6 } f_\gamma = 19 \rightarrow \omega = (19; +\infty)$$

- Принятие решения: $\varphi \notin \omega \rightarrow H_0$

4.2. Фактора B

- Формула вычисления критерия: $F^B = \frac{s_B^2}{s_0^2} = \frac{3750}{150} = 25$

- Схема нахождения критической области:

$$\left. \begin{array}{l} \alpha = 0,05 \\ k_1 = v_B - 1 = 1, \\ k_2 = (v_A - 1) \cdot (v_B - 1) = 2 \end{array} \right\} \text{П6 } f_\gamma = 18,51 \rightarrow \omega = (18,51; +\infty)$$

- Принятие решения: $\varphi \in \omega \rightarrow H_1$

5. Степень влияния фактора на результативный признак (для A и B):

$$\overline{\rho^2}_A = 0; \quad \overline{\rho^2}_B = \frac{625}{891,67} = 0,7 \leq 1$$