Розділ 11

СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗИ ПРО ЗАКОН РОЗПОДІЛУ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

11.1. Перевірка гіпотези про нормальний розподіл

Нехай статистичний розподіл вибірки задано у вигляді послідовності рівновіддалених варіант і відповідних їм частот:

x_i	x_1	x_2	....	x_m
n_i	n_1	n_2	....	n_m

За допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, необхідно:

1) обчислити вибіркове середнє $\bar{x}$ і вибіркове середнє квадратичне відхилення $\sigma_в$;

2) визначити теоретичні частоти

$$n_i' = \frac{n \cdot h}{\sigma_в} \cdot \varphi(u_i), \qquad (11.1)$$

де n – об'єм вибірки; h – крок (різниця між двома сусідніми варіантами); $\varphi(u)$ – диференціальна функція Лапласа; $u_i = \frac{x_i - \bar{x}}{\sigma_в}$;

3) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього:

а) складають розрахункову таблицю

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1.					

і знаходять спостережуване значення критерію Пірсона

$$\chi^2_{\text{спост}} = \sum_{i=1}^{m} \frac{\left(n_i - n'_i\right)^2}{n'_i}; \qquad (11.2)$$

б) за таблицею критичних точок розподілу χ^2 при заданому рівні значущості α і кількості ступенів вільності $k = s - 3$ (s – кількість варіант вибірки) знаходять критичну точку $\chi^2_{\text{крит}}(\alpha; k)$ правосторонньої критичної області. Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, немає підстав відхиляти гіпотезу про нормальний розподіл генеральної сукупності. Іншими словами, емпіричні та теоретичні частоти відрізняються несуттєво (випадково). Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, гіпотезу відхиляють. Іншими словами, емпіричні та теоретичні частоти відрізняються суттєво.

Нехай статистичний розподіл вибірки задано у вигляді послідовності інтервалів і відповідних їм частот

$(x_i;\ x_{i+1}]$	$(x_1;\ x_2]$	$(x_2;\ x_3]$	.	$(x_m;\ x_{m+1}]$
n_i	n_1	n_2	. . . .	n_m

За допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

Правило 2. Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, необхідно:

1) обчислити вибіркове середнє $\overline{x^*}$ і вибіркове середнє квадратичне відхилення σ^*, причому як варіанти x_i^* беруть середнє арифметичне кінців інтервалу:

$$x_i^* = \frac{x_i + x_{i+1}}{2};$$

2) пронумерувати досліджувану випадкову величину X, тобто перейти до випадкової величини $Z = \frac{X - \overline{x^*}}{\sigma^*}$ і обчислити кінці інтервалів: $z_i = \frac{x_i - \overline{x^*}}{\sigma^*}$,

причому найменше значення Z, тобто z_1, вважають рівним $-\infty$, а найбільше, тобто z_{m+1} – рівним ∞;

3) обчислити теоретичні частоти

$$n'_i = n \cdot P_i, \tag{11.3}$$

де n – об'єм вибірки; $P_i = \Phi(z_{i+1}) - \Phi(z_i)$ – ймовірності потрапляння X в інтервали $(x_i;\ x_{i+1}]$; $\Phi(z)$ – функція Лапласа;

4) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього:

А. Складають розрахункову таблицю, за якою знаходять спостережуване значення критерію Пірсона

$$\chi^2_{\text{спост}} = \sum_{i=1}^{m} \frac{\left(n_i - n'_i\right)^2}{n'_i}.$$

Б. За таблицею критичних точок розподілу χ^2 при заданому рівні значущості α і кількості ступенів вільності $k = s - 3$ (s – кількість інтервалів вибірки) знаходять критичну точку $\chi^2_{\text{крит}}(\alpha; k)$ правосторонньої критичної області. Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, немає підстав відхиляти гіпотезу про нормальний розподіл генеральної сукупності. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво (випадково). Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, гіпотезу відхиляють. Інакше кажучи, емпіричні та теоретичні частоти відрізняються суттєво.

Приклад 11.1. Використовуючи критерій Пірсона, при рівні значущості $\alpha = 0{,}05$ перевірити, чи справджується гіпотеза про нормальний розподіл генеральної сукупності X, якщо з цієї сукупності отримано таку вибірку об'ємом $n = 100$:

$(x_i;\ x_{i+1}]$	(3; 7]	(7; 11]	(11; 15]	(15; 19]	(19; 23]	(23; 27]	(27; 31]
n_i	6	16	19	17	15	14	13

Розв'язання. Спочатку обчислимо вибіркове середнє $\overline{x^*}$ і середнє квадратичне відхилення вибірки σ^*. Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	5	9	13	17	21	25	29
n_i	6	16	19	17	15	14	13

Отже,

$$\overline{x^*} = \frac{5\cdot6+9\cdot16+13\cdot19+17\cdot17+21\cdot15+25\cdot14+29\cdot13}{100} = 17{,}52;$$

$$\sigma^* = \sqrt{\frac{5^2\cdot6+9^2\cdot16+13^2\cdot19+17^2\cdot17+21^2\cdot15+25^2\cdot14+29^2\cdot13}{100} - 17{,}52^2} =$$

$$= \sqrt{\frac{35868}{100} - 306{,}95} = \sqrt{51{,}73} \approx 7{,}19;$$

Знайдемо нормовані інтервали $(z_i; z_{i+1}]$, враховуючи, що вибіркове середнє $\overline{x^*} = 17{,}52$ і середнє квадратичне відхилення вибірки $\sigma^* = 7{,}19$. Для цього складемо розрахункову таблицю (лівий кінець першого інтервалу вважаємо рівним $-\infty$, а правий кінець останнього інтервалу – рівним ∞.

i	x_i	x_{i+1}	$x_i - \overline{x^*}$	$x_{i+1} - \overline{x^*}$	$z_i = \frac{x_i - \overline{x^*}}{\sigma^*}$	$z_{i+1} = \frac{x_{i+1} - \overline{x^*}}{\sigma^*}$
1	3	7	-14,52	-10,52	$-\infty$	-1,46
2	7	11	-10,52	-6,52	-1,46	-0,91
3	11	15	-6,52	-2,5	-0,91	-0,35
4	15	19	-2,5	1,48	-0,35	0,21
5	19	23	1,48	5,48	0,21	0,76
6	23	27	5,48	9,48	0,76	1,32
7	27	31	9,48	13,48	1,32	∞

Визначимо теоретичні ймовірності P_i і теоретичні частоти n'_i:

$$n'_i = n \cdot P_i$$

Для цього складемо розрахункову таблицю:

i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n'_i = n \cdot P_i$
1	$-\infty$	-1,46	-0,5000	-0,4279	0,0721	7,21
2	-1,46	-0,91	-0,4279	-0,3186	0,1093	10,93
3	-0,91	-0,35	-0,3186	-0,1368	0,1818	18,18
4	-0,35	0,21	-0,1368	0,0832	0,22	22
5	0,21	0,76	0,0832	0,2764	0,1932	19,32
6	0,76	1,32	0,2764	0,4066	0,1302	13,02
7	1,32	∞	0,4066	0,5000	0,0934	9,34
$\sum$	–	–	–	–	1	100

Порівняємо емпіричні та теоретичні частоти, використовуючи критерій Пірсона. Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$	n_i^2	$\frac{n_i^2}{n_i'}$
1	6	7,21	-1,21	1,4641	0,2031	36	4,9931
2	16	10,93	5,07	25,7049	2,3518	256	23,4218
3	19	18,18	0,82	0,6724	0,0370	361	19,8570
4	17	22	-5	25	1,1364	289	13,1364
5	15	19,32	-4,32	18,6624	0,9660	225	11,6460
6	14	13,02	0,98	0,9604	0,0738	196	15,0538
7	13	9,34	3,66	13,3956	1,4342	169	18,0942
$\sum$	100	100			$\chi^2_{\text{спост}} = 6{,}2023$	6,2023	106,2023

Перевірка:

$$\chi^2_{\text{спост}} = \sum_{i=1}^{m} \frac{n_i^2}{n_i'} - 100.$$

Оскільки $\sum_{i=1}^{m} \frac{n_i^2}{n_i'} - 100 = 106{,}2023 - 100 = 6{,}2023 = \chi^2_{\text{спост}}$, то обчислення проведено правильно.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = s - 3 = 7 - 3 = 4$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}05; 4) = 9{,}49.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, гіпотезу про нормальний розподіл генеральної сукупності не відхиляємо. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво.

11.2. Перевірка гіпотези про рівномірний розподіл

Нехай задано інтервальний статистичний розподіл вибірки:

$(x_i;\ x_{i+1}]$	$(x_1;\ x_2]$	$(x_2;\ x_3]$		$(x_m;\ x_{m+1}]$
n_i	n_1	n_2	....	n_m

n – об'єм вибірки.

За допомогою критерію Пірсона перевірити гіпотезу про рівномірний розподіл генеральної сукупності.

Правило 1. Для того, щоб перевірити гіпотезу про рівномірний розподіл досліджуваної випадкової величини X, тобто про щільність X:

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a;b]; \\ 0, & x \notin [a;b], \end{cases} \tag{11.4}$$

необхідно:

1) оцінити параметри a і b – кінці інтервалу, у якому спостерігалися можливі значення X, за формулами:

$$a^* = \bar{x} - \sqrt{3}\sigma_{\text{в}}, \qquad b^* = \bar{x} + \sqrt{3}\sigma_{\text{в}}, \tag{11.5}$$

де a^* і b^* – оцінки параметрів);

2) знайти щільність ймовірності передбачуваного розподілу:

$$f(x) = \begin{cases} \frac{1}{b^*-a^*}, & x \in [a^*;b^*]; \\ \\ 0, & x \notin [a^*;b^*], \end{cases} \tag{11.6}$$

3) визначити теоретичні частоти:

$$n_1' = n \cdot P_1 = n \cdot \frac{1}{b^*-a^*}(x_1 - a^*); \tag{11.7}$$

$$n_i' = n \cdot P_i = n \cdot \frac{1}{b^*-a^*}(x_i - x_{i-1});$$

$$n_m' = n \cdot P_m = n \cdot \frac{1}{b^*-a^*}(b^* - x_m).$$

4) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 3$, де s – кількість інтервалів вибірки.

Приклад 11.2. Передбачалося, що про стабільність економічної ситуації в країні (відсутність воєн, стихійних лих і т. д.) за останні 50 років можна судити за характером розподілу населення за віком (n_i – млн чол.): при спокійній обстановці розподіл повинен бути рівномірним. В результаті проведеного дослідження, для однієї з країн були отримані такі дані.

$(x_i;\ x_{i+1}]$	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
n_i	14	9	10	8	16	13	12	18

Чи є підстави вважати, що в країні була нестабільна ситуація? При рівні значущості $\alpha = 0{,}01$ перевірити гіпотезу про рівномірний розподіл та зробити відповідні висновки.

Розв'язання. Спочатку обчислимо вибіркове середнє $\overline{x^*}$ і середнє квадратичне відхилення вибірки σ^*. Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	5	15	25	35	45	55	65	75
n_i	14	9	10	8	16	13	12	18

Отже,

$$\overline{x^*} = \frac{5\cdot14+15\cdot9+25\cdot10+35\cdot8+45\cdot16+55\cdot13+65\cdot12+75\cdot18}{100} = \frac{4300}{100} = 43;$$

$$\sigma^* = \sqrt{\frac{5^2\cdot14+15^2\cdot9+25^2\cdot10+35^2\cdot8+45^2\cdot16+55^2\cdot13+65^2\cdot12+75^2\cdot18}{100} - 43^2} =$$

$$= \sqrt{\frac{242100}{100} - 1849} = \sqrt{2421 - 1849} = \sqrt{572} \approx 23{,}92.$$

Оцінимо параметри a і b – кінців інтервалу, у якому спостерігалися можливі значення X:

$$a^* = \bar{x} - \sqrt{3}\sigma_{\text{в}} = 43 - \sqrt{3}\cdot 23{,}92 = 1{,}57;$$

$$b^* = \bar{x} + \sqrt{3}\sigma_{\text{в}} = 43 + \sqrt{3}\cdot 23{,}92 = 84{,}43.$$

Знайдемо щільність ймовірності передбачуваного розподілу:

$$f(x) = \begin{cases} 0{,}0121, & x \in [a^*; b^*]; \\ 0, & x \notin [a^*; b^*]. \end{cases}$$

Визначимо теоретичні частоти:

$$n'_1 = n \cdot P_1 = 100 \cdot 0{,}0121 \cdot (10 - 1{,}57) = 10{,}1;$$
$$n'_2 = n \cdot P_2 = 100 \cdot 0{,}0121(20 - 10) = 12{,}1;$$
$$n'_3 = n \cdot P_3 = 100 \cdot 0{,}0121(30 - 20) = 12{,}1;$$
$$n'_4 = n \cdot P_4 = 100 \cdot 0{,}0121(40 - 30) = 12{,}1;$$
$$n'_5 = n \cdot P_5 = 100 \cdot 0{,}0121(50 - 40) = 12{,}1;$$
$$n'_6 = n \cdot P_6 = 100 \cdot 0{,}0121(60 - 50) = 12{,}1;$$
$$n'_7 = n \cdot P_7 = 100 \cdot 0{,}0121(70 - 60) = 12{,}1;$$
$$n'_8 = n \cdot P_8 = 100 \cdot 0{,}0121(84{,}43 - 70) = 17{,}3.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 3$, де s – кількість інтервалів вибірки.

Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$	n_i^2	$\frac{n_i^2}{n'_i}$
1	14	10,1	3,9	15,21	1,5059	196	19,4059
2	9	12,1	-3,1	9,61	0,7942	81	6,6942
3	10	12,1	-2,1	4,41	0,3645	100	8,2645
4	8	12,1	-4,1	16,81	1,3893	64	5,2893
5	16	12,1	3,9	15,21	1,2570	256	21,157
6	13	12,1	0,9	0,81	0,0669	169	13,9669
7	12	12,1	-0,1	0,01	0,0008	144	11,9008
8	18	17,3	0,7	0,49	0,0283	324	18,7283
$\sum$	100	100			$\chi^2_{\text{спост}} = 5{,}4069$		105,4069

Перевірка:

$$\chi^2_{\text{спост}} = \sum_{i=1}^{m} \frac{n_i^2}{n_i'} - 100.$$

Оскільки $\sum_{i=1}^{m} \frac{n_i^2}{n_i'} - 100 = 105{,}4069 - 100 = 5{,}4069 = \chi^2_{\text{спост}}$, то обчислення проведено правильно.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}01$ і кількості ступенів вільності

$$k = s - 3 = 8 - 3 = 5$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}01; 5) = 15{,}09.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, гіпотезу про рівномірний розподіл генеральної сукупності не відхиляємо. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво. Отже, підстави вважати, що в країні була нестабільна ситуація, немає.

11.3. Перевірка гіпотези про показниковий розподіл

Нехай задано інтервальний статистичний розподіл вибірки

$(x_i;\ x_{i+1}]$	$(x_1;\ x_2]$	$(x_2;\ x_3]$		$(x_m;\ x_{m+1}]$
n_i	n_1	n_2	....	n_m

n – об'єм вибірки.

За допомогою критерію Пірсона перевірити гіпотезу про показниковий розподіл генеральної сукупності.

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про показниковий розподіл досліджуваної випадкової величини X, необхідно:

1) знайти вибіркове середнє $\bar{x}$, узявши як варіанти середини інтервалів:

$$x_i^* = \frac{x_i + x_{i+1}}{2};$$

2) за оцінку параметра λ показникового розподілу взяти величину, обернену до вибіркового середнього,

$$\lambda^* = \frac{1}{\bar{x}}, \tag{11.8}$$

3) знайти ймовірності потрапляння випадкової величини X у частинні інтервали $(x_i;\ x_{i+1}]$ за формулою

$$P_i = P(x_i < X < x_{i+1}) = e^{-\lambda^* x_i} - e^{-\lambda^* x_{i+1}}; \tag{11.9}$$

4) обчислити теоретичні частоти

$$n_i' = n \cdot P_i;$$

5) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, узявши кількість ступенів вільності $k = s - 2$, де s – кількість інтервалів вибірки.

Приклад 11.3. В результаті випробування 200 елементів на тривалість роботи отримано емпіричне розподіл, наведене в таблиці (у першому рядку вказані інтервали часу в годинах, в другому рядку – частоти, тобто кількість елементів, які працювали в межах відповідного інтервалу). Використовуючи критерій Пірсона, при рівні значущості $\alpha = 0{,}05$ перевірити, чи узгоджується гіпотеза про показниковий розподіл генеральної сукупності X з емпіричним розподілом вибірки.

$(x_i;\ x_{i+1}]$	$(0;\ 5]$	$(5;\ 10]$	$(10;\ 15]$	$(15;\ 20]$	$(20;\ 25]$	$(25;\ 30]$
n_i	133	45	15	4	2	1

Розв'язання. Спочатку обчислимо вибіркове середнє $\bar{x}$. Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	2,5	7,5	12,5	17,5	22,5	27,5
n_i	133	45	15	4	2	1

Отже,

$$\bar{x} = \frac{2{,}5\cdot 133 + 7{,}5\cdot 45 + 12{,}5\cdot 15 + 17{,}5\cdot 4 + 22{,}5\cdot 2 + 27{,}5\cdot 1}{200} = \frac{1000}{200} = 5.$$

Оцінимо параметр λ показникового розподілу взявши величину, обернену до вибіркового середнього,

$$\lambda^* = \frac{1}{\bar{x}} = \frac{1}{5} = 0{,}2.$$

Знайдемо ймовірності потрапляння випадкової величини X у частинні інтервали $(x_i;\ x_{i+1}]$ за формулою

$$P_1 = P(0 < X < 5) = e^{-0{,}2\cdot 0} - e^{-0{,}2\cdot 5} =$$
$$= 1 - e^{-1} = 1 - 0{,}3679 = 0{,}6321;$$

$$P_2 = P(5 < X < 10) = e^{-0{,}2\cdot 5} - e^{-0{,}2\cdot 10} =$$
$$= e^{-1} - e^{-2} = 0{,}3679 - -0{,}1353 = 0{,}2326;$$

$$P_3 = P(10 < X < 15) = e^{-0{,}2\cdot 10} - e^{-0{,}2\cdot 15} =$$
$$= e^{-2} - e^{-3} = 0{,}1353 - -0{,}0498 = 0{,}0855;$$

$$P_4 = P(15 < X < 20) = e^{-0{,}2\cdot 15} - e^{-0{,}2\cdot 20} =$$
$$= e^{-3} - e^{-4} = 0{,}0498 - -0{,}0183 = 0{,}0315;$$

$$P_5 = P(20 < X < 25) = e^{-0{,}2\cdot 20} - e^{-0{,}2\cdot 25} =$$
$$= e^{-4} - e^{-5} = 0{,}0183 - -0{,}0067 = 0{,}0116;$$

$$P_6 = P(25 < X < 30) = e^{-0{,}2\cdot 25} - e^{-0{,}2\cdot 30} =$$
$$= e^{-5} - e^{-6} = 0{,}0067 - -0{,}0025 = 0{,}0042.$$

Визначимо теоретичні частоти:

$n_1' = n \cdot P_1 = 200 \cdot 0{,}6321 = 126{,}42;$

$n_2' = n \cdot P_2 = 200 \cdot 0{,}2326 = 46{,}52;$

$n_3' = n \cdot P_3 = 200 \cdot 0{,}0855 = 17{,}1;$

$n_4' = n \cdot P_4 = 200 \cdot 0{,}0315 = 6{,}3;$

$n_5' = n \cdot P_5 = 200 \cdot 0{,}0116 = 2{,}32;$

$n_6' = n \cdot P_6 = 200 \cdot 0{,}0042 = 0{,}84.$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 2$, де s – кількість інтервалів вибірки.

Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
1	133	126,42	6,58	43,2964	0,3425
2	45	46,52	-1,52	2,3104	0,0497
3	15	17,1	-2,1	4,41	0,2579
4	4	6,3	-2,3	5,29	0,8397
5	2	2,32	-0,32	0,1024	0,0441
6	1	0,84	0,16	0,0256	0,0305
$\sum$					$\chi^2_{\text{спост}} = 1{,}5644$

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = s - 3 = 6 - 2 = 1$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}05; 3) = 9{,}49.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, гіпотезу про показниковий розподіл генеральної сукупності не відхиляємо. Інакше кажучи, емпіричні та теоретичні частоти відрізняються несуттєво.

11.4. Перевірка гіпотези про біноміальний розподіл

Нехай проведено n дослідів. Кожний дослід складається із N незалежних випробувань, для яких ймовірність появи деякої події A одна й та ж. Реєструється кількість появ події A в кожному досліді. У підсумку отримаємо статистичний розподіл дискретної випадкової величини X, яка характеризує кількість появ події A (у першому рядку наведено кількість появ події A в одному досліді, а в другому рядку – частоту n_i, тобто кількість дослідів, у яких зареєстровано x_i появ події A):

x_i	0	1	....	N
n_i	n_1	n_2		n_N

За допомогою критерію Пірсона перевірити гіпотезу про біноміальний закон розподілу дискретної випадкової величини X.

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про біноміальний розподіл дискретної випадкової величини X, необхідно:

1) знайти за формулою Бернуллі

$$P_N(i) = C_N^i p^i (1-p)^{N-i} \tag{11.10}$$

ймовірності $P_N(i)$ появи події A рівно i разів у N випробуваннях ($i = 0,1,2,\dots,s$), де s – максимальна кількість спостережуваних появ події A в одному досліді, тобто $s \le N$);

2) знайти теоретичні частоти

$$n_i' = n \cdot P_N(i);$$

де n – кількість дослідів;

3) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, поклавши кількість ступенів вільності $k = s - 1$ (вважається, що ймовірність p появи події A задана, тобто вона не оцінювалась за вибіркою і не об'єднувалися малочисельні частоти).

Якщо ймовірність p була оцінена за вибіркою, то $k = s - 2$. Якщо, крім того, було об'єднано малочисельні частоти, то s – кількість варіант вибірки, які залишилися після об'єднання частот.

Приклад 11.4. Чотири монети були підкинуті 20160 разів, водночас комбінації: чотири «герби», три «герби» і «цифра», два «герби» і дві «цифри», один «герб» і три «цифри», чотири «цифри» з'явилися таку відповідну кількість разів:

1181, 4909, 7583, 5085, 1402.

Чи свідчать ці дані про те, що кількість «гербів», яка з'явилась на чотирьох монетах має біноміальний розподіл. Застосувати критерій χ^2 , при рівні значущості $\alpha = 0{,}05$.

Розв'язання. Спочатку знайдемо за формулою Бернуллі (11.10) ймовірності $P_N(i)$ появи події A – кількість випавши гербів. Дані запишемо у таблицю.

$$P_4(4) = C_4^4 \cdot 0{,}5^4 \cdot (1 - 0{,}5)^{4-4} = 1 \cdot 0{,}0625 \cdot 1 = 0{,}0625;$$

$$P_4(3) = C_4^3 \cdot 0{,}5^3 \cdot \left(1 - 0{,}5\right)^{4-3} = 4 \cdot 0{,}125 \cdot 0{,}5 = 0{,}25;$$

$$P_4(2) = C_4^2 \cdot 0{,}5^2 \cdot (1 - 0{,}5)^{4-2} = 6 \cdot 0{,}25 \cdot 0{,}25 = 0{,}375;$$

$$P_4(1) = C_4^1 \cdot 0{,}5^1 \cdot (1 - 0{,}5)^{4-1} = 4 \cdot 0{,}5 \cdot 0{,}125 = 0{,}25;$$

$$P_4(0) = C_4^0 \cdot 0{,}5^0 \cdot (1 - 0{,}5)^{4-0} = 1 \cdot 1 \cdot 0{,}0625 = 0{,}0625.$$

X	4	3	2	1	0
$P(X)$	0,0625	0,25	0,375	0,25	0,0625

Знайдемо теоретичні частоти, використовуючи формулу:

$$n_i' = n \cdot P_N(i).$$

$n_1' = 20160 * 0{,}0625 = 1260;$

$n_2' = 20160 * 0{,}25 = 5040;$

$n_3' = 20160 * 0{,}375 = 7560;$

$n_4' = 20160 * 0{,}25 = 5040;$

$n_5' = 20160 * 0{,}0625 = 1260.$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього складемо розрахункову таблицю:

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
0	1181	1260	-79	6241	4,9532
1	4909	5040	-131	17161	3,405
2	7583	7560	23	529	0,0699
3	5085	5040	45	2025	0,4018

4	1402	1260	142	20164	16,0032
5	1181	1260	-79	6241	4,9532
$\sum$					$\chi^2_{\text{спост}} = 24{,}8331$

Із розрахункової таблиці отримаємо

$$\chi^2_{\text{спост}} = 24{,}8331.$$

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = s - 2 = 5 - 2 = 3$$

знаходимо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}05; 3) = 7{,}82.$$

Оскільки $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, є підстави відхиляти гіпотезу про розподіл випадкової величини X (кількості «гербів» за біноміальним законом).

11.5. Перевірка гіпотези про розподіл Пуассона

Нехай задано точковий статистичний розподіл вибірки. За допомогою критерію Пірсона перевірити гіпотезу про розподіл генеральної сукупності за законом Пуассона.

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про те, що досліджувана випадкова величина X розподілена за законом Пуассона, необхідно:

1) знайти за цим статистичним розподілом вибіркове середнє $\bar{x}$;

2) взяти за оцінку параметра λ розподілу Пуассона вибіркове середнє

$$\lambda^* = \bar{x}; \qquad (11.11)$$

3) знайти за формулою Пуассона (або за таблицею значень функції Пуассона) ймовірності P_i появи рівно i подій у n випробуваннях

$(i = 1, 2, 3, \ldots, r)$, де r – максимальна кількість спостережуваних подій; n – об'єм вибірки;

4) визначити теоретичні частоти

$$n_i' = n \cdot P_i;$$

5) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, узявши кількість ступенів вільності $k = s - 2$, де s – кількість варіант вибірки (якщо проводилося об'єднання малочисельних частот в одну групу, то s – кількість варіант, які залишилися після об'єднання).

Приклад 11.4. У $n = 1000$ перевірках партій товару реєструвалася кількість x_i неякісної продукції, внаслідок чого було отримано такий статистичний розподіл кількості браку x_i в n_i партіях товару:

x_i	0	1	2	3	4	5	6	7
n_i	242	349	234	107	43	21	3	1

При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про те, що кількість неякісної продукції X розподілена за законом Пуассона.

Розв'язання. Спочатку знайдемо вибіркове середнє:

$$\bar{x} = \frac{0 \cdot 242 + 1 \cdot 349 + 2 \cdot 234 + 3 \cdot 107 + 4 \cdot 43 + 5 \cdot 21 + 6 \cdot 3 + 7 \cdot 1}{1000} = \frac{1440}{1000} = 1{,}44.$$

Візьмемо за оцінку параметра λ розподілу Пуассона вибіркове середнє

$$\lambda^* = \bar{x} = 1{,}44.$$

Припускаємо, що закон Пуассона має вигляд:

$$P_{1000}(i) = 1{,}44^i \cdot \frac{e^{-1{,}44}}{i!}.$$

Поклавши $i = 0, 1, 2, \ldots, 7$, обчислимо ймовірності $P_i = P_{1000}(i)$:

$$P_0 = 1{,}44^0 \cdot \frac{e^{-1{,}44}}{0!} \approx 0{,}2369; \qquad P_1 = 1{,}44^1 \cdot \frac{e^{-1{,}44}}{1!} \approx 0{,}3412;$$

$$P_2 = 1{,}44^2 \cdot \frac{e^{-1{,}44}}{2!} \approx 0{,}2456; \qquad P_3 = 1{,}44^3 \cdot \frac{e^{-1{,}44}}{3!} \approx 0{,}1179;$$

$P_4 = 1{,}44^4 \cdot \frac{e^{-1{,}44}}{4!} \approx 0{,}0424;$ $P_5 = 1{,}44^5 \cdot \frac{e^{-1{,}44}}{5!} \approx 0{,}0122;$

$P_6 = 1{,}44^6 \cdot \frac{e^{-1{,}44}}{6!} \approx 0{,}0029;$ $P_7 = 1{,}44^7 \cdot \frac{e^{-1{,}44}}{7!} \approx 0{,}0006;$

Знайдемо теоретичні частоти $n'_i = n \cdot P_i$:

$n'_0 = 1000 * 0{,}2369 = 236{,}9;$ $n'_1 = 1000 * 0{,}3112 = 311{,}2;$

$n'_2 = 1000 * 0{,}2369 = 236{,}9;$ $n'_3 = 1000 * 0{,}1179 = 117{,}9;$
$n'_4 = 1000 * 0{,}0424 = 42{,}4;$ $n'_5 = 1000 * 0{,}0122 = 12{,}2;$

$n'_6 = 1000 * 0{,}0029 = 2{,}9;$ $n'_7 = 1000 * 0{,}0006 = 0{,}6.$

Оскільки частоти $n_6 = 3$ і $n_7 = 1$ – малочисельні (менше п'яти), об'єднаємо їх із частотою $n_5 = 21$, а саме

$$n_5 = 21 + 3 + 1 = 25.$$

Як теоретичну частоту, що відповідає об'єднаній частоті, візьмемо суму відповідних теоретичних частот:

$$n'_5 = 12{,}2 + 2{,}9 + 0{,}6 = 15{,}7.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього складемо розрахункову таблицю:

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
0	242	236,9	5,1	26,01	0,1098
1	349	341,2	7,8	60,84	0,1784
2	234	245,6	-11,6	134,56	0,5479
3	107	117,9	-10,9	118,81	1,0077
4	43	42,4	0,6	0,36	0,0085
5	25	15,7	9,3	86,49	5,5089
$\sum$	1000				$\chi^2_{\text{спост}} = 7{,}3612$

Із розрахункової таблиці отримаємо

$$\chi^2_{\text{спост}} = 7{,}3612.$$