Лекція 12. Перевірка правильності непараметричних гіпотез

Із усієї множини критеріїв згоди, що використовуються під час перевірки гіпотез стосовно розподілів, частіше від інших (є найпоширенішим) застосовується критерій χ^2 Пірсона.

12.1. Критерій χ^2 Пірсона

Щоб при заданому рівні значущості α перевірити основну гіпотезу H_0: генеральна сукупність розподілена за відповідним законом розподілу, потрібно:

1) Обчислити теоретичні частоти n_i' ($n_i' = np_i$) для варіант вибірки за певними формулами у залежності від виду запропонованого розподілу. При цьому вид передбачуваного розподілу визначається за характером отриманої гістограми. Для цього порівнюють вид гістограми з графіками щільності відомих розподілів:

Якщо побудована гістограма виявиться несхожою ні на одне з наведених розподілів, то це означає, що потрібно брати інші закони розподілу. Їх досить багато, і відповідні графіки можна знайти у підручниках або довідниках по теорії ймовірностей.

2) Обчислити (згідно з обраною гіпотезою про закон розподілу) спостережене значення критерію $\chi^2 = \sum_{i=1}^{r} \frac{\left(n_i - n_i'\right)^2}{n_i'}$. Для наближеного обчислення спостереженого χ^2_{cn} зазвичай використовують наступну таблицю:

<table>
<tr><td colspan="2">x</td><td rowspan="2">n_i</td><td rowspan="2">$p_i = P_n\left(x_i\right)$</td><td rowspan="2">$n_i' = np_i$</td><td rowspan="2">$\left(n_i - n_i'\right)^2$</td><td rowspan="2">$\frac{\left(n_i - n_i'\right)^2}{n_i'}$</td></tr>
<tr><td>x_i</td><td>x_{i+1}</td></tr>
<tr><td colspan="2">$\sum$</td><td>$= n$</td><td>$=1$</td><td>$= n$</td><td></td><td>$= \chi^2 \Rightarrow \chi^2_{спост}$</td></tr>
</table>

3) Знайти степінь свободи (k). У кожному розподілі у формулі щільності розподілу ймовірностей $f(x)$ присутні параметри розподілу (l).

Зазвичай оцінюють по вибірці математичне сподівання (a або m), дисперсію (D) і стандартне відхилення σ, при цьому використовують відповідно вибіркове середнє $\overline{x}_B$, виправлену вибіркову дисперсію S^2 і виправлене вибіркове стандартне відхилення S, якщо $n < 30$. Для будь-якого з розподілів числові характеристики виражаються через параметри розподілу відомими співвідношеннями:

За формулами передбачуваного теоретичного закону розподілу підраховують теоретичні частоти. Для цього потрібно підрахувати ймовірності влучень у кожен з інтервалів з використанням формул передбачуваного розподілу.

У будь-якому випадку ймовірність попадання у інтервал підраховується як різниця значень деякої функції на кінцях цього інтервалу.

4) Знайти з таблиці критичну точку χ^2_{kp}, яка відповідає заданому рівню значущості α та степені свободи k;

5) Порівняти $\chi^2_{спост}$ та χ^2_{kp} зробити висновок: якщо $\chi^2_{спост} < \chi^2_{kp}$, то немає підстав відхилити гіпотезу H_0; якщо $\chi^2_{спост} \geq \chi^2_{kp}$, то гіпотезу H_0 треба відхилити.

Рекомендації по використанню критерію:

1. Обсяг вибірки повинен бути достатньо великий в усякому разі, не менше 50. Оскільки статистика

$$\chi^2 = \sum_{i=1}^{r} \frac{\left(n_i - n_i'\right)^2}{n_i'} \tag{12.1}$$

має χ^2 - розподіл тільки $n \to \infty$, тому необхідно, щоб у кожному інтервалі була достатня кількість спостережень. Кожна група повинна містити не менше 5-8 варіант. Якщо у деяких інтервалах частоти менші за п’ять, слід об’єднати кілька сусідніх інтервалів.

2. Оскільки можливі помилки першого і другого роду, особливо якщо узгодження теоретичних і емпіричних частот «дуже добре», слід проявляти обережність. Наприклад, можна повторити дослід, збільшивши число спостережень, скористатися іншими критеріями, побудувати графік розподілу.

3. Для спрощення обчислень формулу для $\chi^2_{спост}$ перетворюють до вигляду $\chi^2_{спост} = \left(\sum_{i=1}^{r} \frac{n_i^2}{n_i'} \right) - n$.

Зауваження. Якщо усі емпіричні частоти збігаються з теоретичними, то маємо $\chi^2 = 0$, у протилежному випадку — завжди $\chi^2 > 0$. Тому критична область є правобічною.

12.2. Алгоритми перевірок гіпотез для неперервних випадкових величин

*12.2.1. **Перевірка гіпотези про нормальний розподіл генеральної сукупності***

Нехай гістограма має вигляд, зображений на рис. 12.1. Можемо припустити, що ознака генеральної сукупності розподілена за нормальним законом.

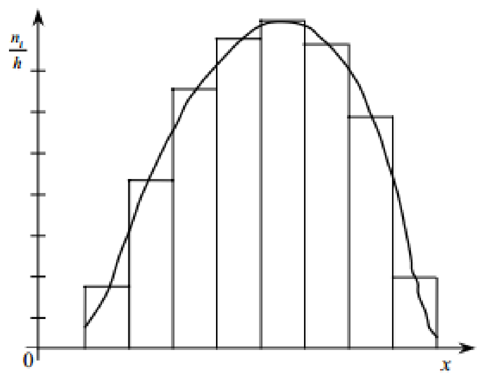


Рис. 12.1

Для нормального закону розподілу теоретичні частоти обчислюють за такими формулами:

$$n_i' = np_i = \frac{nh}{\sigma_B}\varphi\left(\frac{x_i - \overline{x}_B}{\sigma_B}\right), \quad h = x_{i-1} - x_i, \tag{12.2}$$

де $\varphi(x)$ - функція Гаусса; h - довжина інтервалу;

$$n_i' = np_i = n\left(\Phi\left(\frac{x_i - \overline{x}_B}{\sigma_B}\right) - \Phi\left(\frac{x_{i-1} - \overline{x}_B}{\sigma_B}\right)\right), \tag{12.3}$$

де $\Phi(x)$ - функція Лапласа.

Формула (12.2) використовується, якщо результати вибірки подано у вигляді точкового статистичного ряду з рівновіддаленими варіантами, формула (12.3) використовується, якщо результати вибірки подано у вигляді інтервального статистичного ряду.

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

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

2) обчислюють теоретичні частоти за формулою $n_i' = np_i = n\left(\Phi\left(\frac{x_{i+1} - \overline{x}_B}{\sigma_B}\right) - \Phi\left(\frac{x_i - \overline{x}_B}{\sigma_B}\right)\right)$, причому, так як для нормального розподілу $-\infty < x < +\infty$, то вважаємо, що $x_1 = -\infty, x_r = +\infty$.

3) обчислюють спостережуване значення критерію $\chi^2_{спост}$. за формулою $\chi^2_{спост} = \sum_{i=1}^{r}\frac{(n_i - np_i)^2}{np_i}$.

4) визначається критична точка $\chi^2{}_{кр}$ (додаток 8) правобічної критичної області за заданого рівня значущості α і числом ступенів вільності $k = r - 3$ (оскільки число параметрів у нормально розподіленого закону $l = 2$, r - число інтервалів розподілу, причому, якщо в інтервальному ряду частот варіант менше п'яти $(n_i < 5)$, тоді ряд і відповідні їм теоретичні частоти об'єднуємо з

попередніми чи наступними значеннями. При цьому кількість інтервалів вибірки зменшиться).

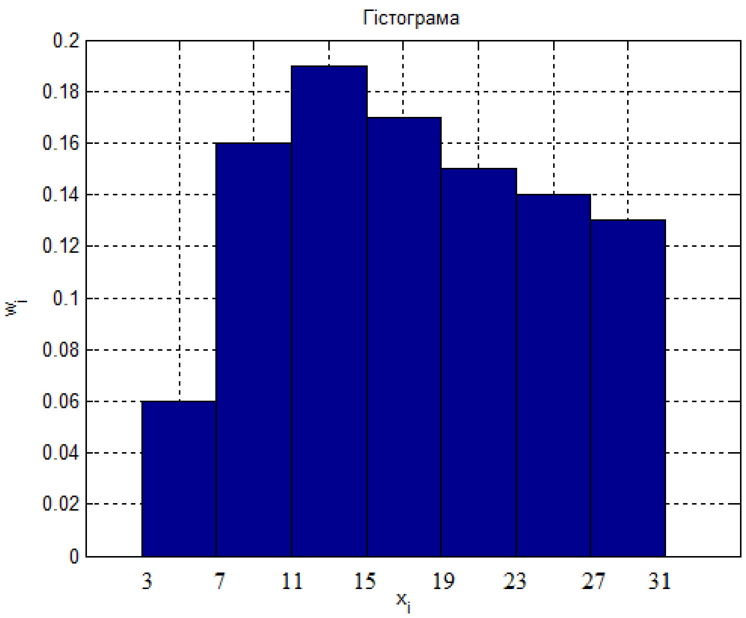
5) зіставляємо $\chi^2_{кр}$ і $\chi^2_{спост}$. Якщо $\chi^2_{кр} > \chi^2_{спост}$, то немає підстав, щоб відхилити нульову гіпотезу; якщо ж $\chi^2_{кр} \le \chi^2_{спост}$, то нульову гіпотезу відхиляємо.

Приклад 12.1. Використовуючи критерій Пірсона, при рівні значущості $\alpha = 0,05$ перевірити, чи справджується гіпотеза про висунутий закон розподілу генеральної сукупності, 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:

Побудуємо діаграму відносних частот інтервального ряду та висунимо припущення про закон розподіл генеральної сукупності.



Згідно гістограми можна висунути, що генеральна сукупність розподілена за нормальним законом розподілу.

Перевіримо, чи справджується гіпотеза про нормальний розподіл генеральної сукупності.

Для цього перетворимо інтервальний статистичний розподіл вибірки на дискретний:

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

$n=100;\quad \overline{x}_B=17,52;\quad \sigma_B=7,19$.

Знайдемо нормовані інтервали $(z_i; z_{i+1})$, враховуючи вибіркове середнє і середнє квадратичне відхилення вибірки. Для цього складемо розрахункову таблицю (лівий кінець першого інтервалу вважаємо рівним $-\infty$, а правий кінець останнього інтервалу рівним $+\infty$.

i	x_i	x_{i+1}	$z_i=\dfrac{x_i-\overline{x}_B}{\sigma_B}$	$z_{i+1}=\dfrac{x_{i+1}-\overline{x}_B}{\sigma_B}$	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i=\Phi(z_{i+1})-\Phi(z_i)$
1	3	7	$-\infty$	-1,46	-0,5	-0,4279	0,0721
2	7	11	-1,46	-0,91	-0,4279	-0,3186	0,1093
3	11	15	-0,91	-0,35	-0,3186	-0,1368	0,1818
4	15	19	-0,35	0,21	-0,1368	0,0832	0,22
5	19	23	0,21	0,76	0,0832	0,2764	0,1932
6	23	27	0,76	1,32	0,2764	0,4066	0,1302
7	27	31	1,32	$+\infty$	0,4066	0,5	0,0934
Сума							**1**

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'	$(n_i-n_i')^2$	$\dfrac{(n_i-n_i')^2}{n_i'}$
1	6	7,21	-1,21	1,4641	0,2031
2	16	10,93	5,07	25,7049	2,3518
3	19	18,18	0,82	0,6724	0,0370
4	17	22	-5	25	1,1364
5	15	19,32	-4,32	18,6624	0,9660
6	14	13,02	0,98	0,9604	0,0738
7	13	9,34	3,66	13,3956	1,4342
$\sum$	**100**	**100**	**-**	**-**	**6,2023**

Отже $\chi^2_{спост}=6,2023$.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності $k = r - l - 1 = 7 - 2 - 1 = 4$ знайдемо критичну точку правосторонньої критичної області (додаток 8):

$$\chi^2_{кр}(\alpha; k) = \chi^2_{кр}(0{,}05; 4) = 9{,}5.$$

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

12.2.2. Перевірка гіпотези про рівномірний розподіл

Якщо гістограма має вигляд, зображений на рис. 12.2, можемо висунути гіпотезу про рівномірний закон розподілу сукупності.

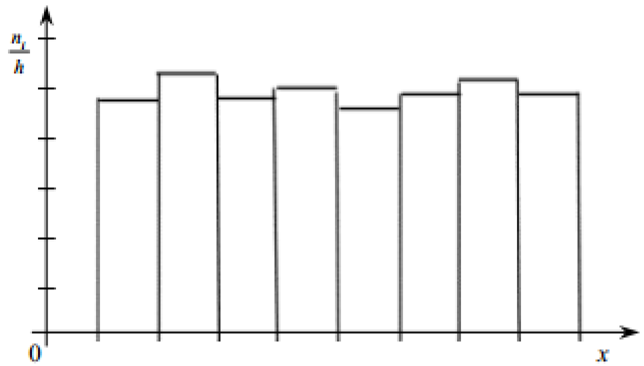


Рис. 12.2.

Для цього потрібно оцінити параметри a і b – кінці інтервалу, у якому спостерігалися можливі значення X, за формулами: $a^* = \bar{x}_B - \sqrt{3}\sigma_B$ та $b^* = \bar{x}_B + \sqrt{3}\sigma_B$, де a^* і b^* - оцінки параметрів. Після цього обчислимо ймовірності p_i — імовірності потрапляння випадкової величини X у частинні інтервали, тобто $p_1 = \dfrac{x_1 - a^*}{b^* - a^*}$; $p_i = \dfrac{x_{i+1} - x_i}{b^* - a^*}$, де $i = \overline{2, r-1}$; $p_r = \dfrac{b^* - x_{r-1}}{b^* - a^*}$ та визначимо теоретичні частоти n_i' за формулою: $n_i' = np_i$, де $n = \sum_{i=1}^{r} n_i$ — обсяг вибірки, і знайдемо спостережуване значення критерію Пірсона $\chi^2_{кр}$.

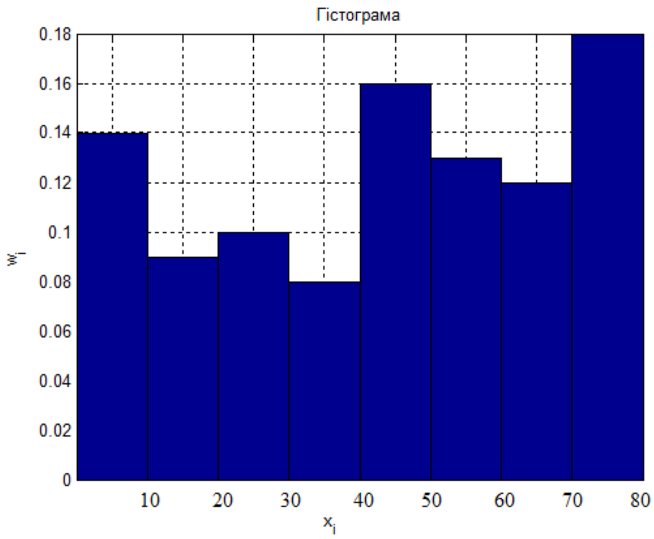
Визначаємо $\chi^2_{кр}$ за заданим рівнем значущості α та числом ступенів вільності $k = r - l - 1$, де $l = 2$ (параметри a^* і b^*) і приймаємо одне з наступних рішень:

• якщо $\chi^2_{спост} < \chi^2_{кр}$, то вважаємо, що немає достатніх підстав для відхилення гіпотези H_0 про рівномірний розподіл випадкової величини X на відрізку $[a;b]$;

• якщо $\chi^2_{спост} \geq \chi^2_{кр}$, то гіпотезу H_0 відхиляємо.

Приклад 12.2. 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	19	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 $\overline{x}_B$ і середнє квадратичне відхилення вибірки σ_B. Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i^*	5	15	25	35	45	55	65	75
n_i	14	19	10	8	16	13	12	18

Отже, $n=100$, $\overline{x}_B=43$, $\sigma_B=23{,}92$.

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

$a^*=\overline{x}_B-\sqrt{3}\sigma_B=43-\sqrt{3}\cdot 23{,}92=1{,}57$;

$b^*=\overline{x}_B+\sqrt{3}\sigma_B=43+\sqrt{3}\cdot 23{,}92=84{,}43$.

Знайдемо щільність ймовірності передбачуваного розподілу:

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

Визначимо теоретичні частоти:

$n_1^{'}=n\cdot p_1=100\cdot 0{,}0121\cdot(10-1{,}57)=10{,}1$

$n_2^{'}=n\cdot p_2=100\cdot 0{,}0121\cdot(20-10)=12{,}1$

$n_3^{'}=n\cdot p_3=100\cdot 0{,}0121\cdot(30-20)=12{,}1$

$n_4^{'}=n\cdot p_4=100\cdot 0{,}0121\cdot(40-30)=12{,}1$

$n_5^{'}=n\cdot p_5=100\cdot 0{,}0121\cdot(50-40)=12{,}1$

$n_6^{'}=n\cdot p_6=100\cdot 0{,}0121\cdot(60-50)=12{,}1$

$n_7^{'}=n\cdot p_7=100\cdot 0{,}0121\cdot(70-60)=12{,}1$

$n_8^{'}=n\cdot p_8=100\cdot 0{,}0121\cdot(84{,}43-70)=17{,}3.$

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

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

i	n_i	$n_i^{'}$	$n_i - n_i^{'}$	$\left(n_i - n_i^{'}\right)^2$	$\frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}$
1	14	10,1	3,9	15,21	1,5059
2	9	12,1	-3,1	9,61	0,7942
3	10	12,1	-2,1	4,41	0,3645
4	8	12,1	-4,1	16,81	1,3893
5	16	12,1	3,9	15,21	1,2570
6	13	12,1	0,9	0,81	0,0669
7	12	12,1	-0,1	0,01	0,0008
8	18	17,3	0,7	0,49	0,0283
$\sum$	**100**	**100**	**-**	**-**	**5,4069**

Отже $\chi^2_{спост} = 5,4069$.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,01$ і кількості ступенів вільності $k = r - l - 1 = 8 - 2 - 1 = 5$ знайдемо критичну точку правосторонньої критичної області (додаток 8):

$$\chi^2_{кр}\left(\alpha; k\right) = \chi^2_{кр}\left(0,01; 5\right) = 15,1.$$

Оскільки $\chi^2_{спост} < \chi^2_{кр}$, 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.

12.2.3. Перевірка гіпотези про розподіл неперервної випадкової величини за показниковим законом

Якщо гістограма має вигляд, зображений на рис. 12.3, маємо підставу вважати, що ознака розподілена за показниковим законом, оскільки лінія, що з’єднує середини сторін прямокутників, подібна до графіка щільності показникового закону.

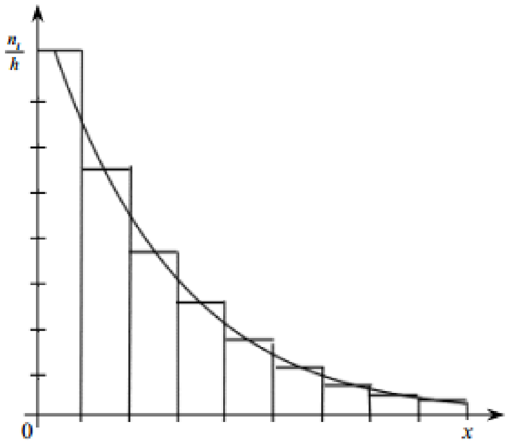


Рис. 12.3

Показниковий закон розподілу задається щільністю розподілу ймовірностей, тобто диференціальною функцією:

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & для \quad x\geq 0,\\ 0 & для \quad x<0,\end{cases}$$ де $\lambda>0$ - стала.

Як і у попередніх розподілах, вихідні вибіркові дані групуємо, потім подаємо у вигляді послідовності частинних інтервалів і відповідних їм частот, а далі обчислюємо вибіркову середню $\overline{x}_B$ (якщо $\overline{x}_B=\sigma_B$ — це ознака показникового розподілу).

За оцінку параметра λ показникового розподілу беремо величину, обернену до $\overline{x}_B$: $\lambda=\dfrac{1}{\overline{x}_B}$.

Імовірності потрапляння випадкової величини X у частинні інтервали x_i, x_{i+1}, тобто $p_i=P\left(x_i<x<x_{i+1}\right)$ знаходимо за формулою:

$$p_i=e^{\frac{-x_i}{\overline{x}_B}}-e^{\frac{-x_{i+1}}{\overline{x}_B}}=e^{-\lambda\cdot x_i}-e^{-\lambda\cdot x_{i+1}},$$

причому, оскільки $0<x<\infty$, то приймаємо $x_1=0$; $x_r=\infty$ (останнє значення x).

Якщо параметр λ заданий за умовою, то його оцінка замінюється точним значенням.

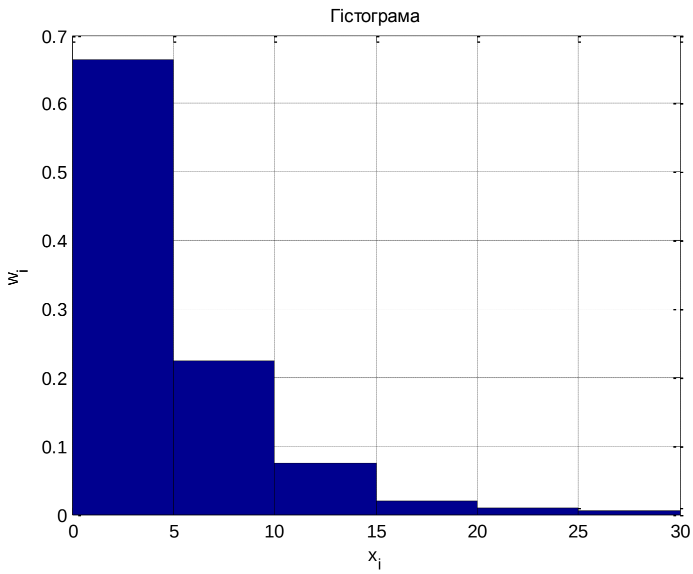
Критичне значення χ^2 визначаємо за заданим рівнем значущості та числом ступенів вільності $k = r - l - 1$, де l дорівнює 1 або 0 залежно від того, оцінювався чи ні за вибіркою параметр λ.

У результаті приймаємо одне з рішень: якщо $\chi^2_{спост} < \chi^2_{кр}$, то вважаємо, що немає достатніх підстав відхилити нульову гіпотезу про розподіл випадкової величини X за показниковим законом; якщо ж $\chi^2_{спост} > \chi^2_{кр}$, то нульову гіпотезу відхиляємо.

Приклад 12.3. 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: Побудуємо гістограму відносних частот



Спочатку обчислимо вибіркове середнє $\overline{x}_B$. 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, $n = 200$; $\overline{x}_B = 5$.

Оцінимо параметр λ показникового розподілу взявши величину, обернену до вибіркового середнього, $\lambda = \frac{1}{\overline{x}_B} = \frac{1}{5} = 0,2$.

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

$$p_1 = P\left(0 < X < 5\right) = e^{-0,2\cdot 0} - e^{-0,2\cdot 5} = 1 - e^{-1} = 1 - 0,3679 = 0,6321$$

$$p_2 = P\left(5 < X < 10\right) = e^{-0,2\cdot 5} - e^{-0,2\cdot 10} = e^{-1} - e^{-2} = 0,3679 - 0,1353 = 0,2326$$

$$p_3 = P\left(10 < X < 15\right) = e^{-0,2\cdot 10} - e^{-0,2\cdot 15} = e^{-2} - e^{-3} = 0,1353 - 0,0498 = 0,0855$$

$$p_4 = P\left(15 < X < 20\right) = e^{-0,2\cdot 15} - e^{-0,2\cdot 20} = e^{-3} - e^{-4} = 0,0498 - 0,0183 = 0,0315$$

$$p_5 = P\left(20 < X < 25\right) = e^{-0,2\cdot 20} - e^{-0,2\cdot 25} = e^{-4} - e^{-5} = 0,0183 - 0,0067 = 0,0116$$

$$p_6 = P\left(25 < X < \infty\right) = e^{-0,2\cdot 25} - e^{-0,2\cdot \infty} = e^{-5} - e^{\infty} = 0,0067 - 0 = 0,0067$$

Визначимо теоретичні частоти:

$$n_1^{'} = n \cdot p_1 = 200 \cdot 0,6321 = 126,42$$

$$n_2^{'} = n \cdot p_2 = 200 \cdot 0,2326 = 46,52$$

$$n_3^{'} = n \cdot p_3 = 200 \cdot 0,0855 = 17,1$$

$$n_4^{'} = n \cdot \left(p_4 + p_5 + p_6\right) = 200 \cdot 0,0498 = 9,96.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = r - l - 1$, де r – кількість інтервалів вибірки, $l = 1$ - параметр $\left(\lambda\right)$.

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

i	n_i	$n_i^{'}$	$n_i - n_i^{'}$	$\left(n_i - n_i^{'}\right)^2$	$\frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}$
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	7	9,96	-2,96	8,7616	0,8797
$\sum$	**200**				**1,5298**

Отже $\chi^2_{спост} = 1,5298$.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,05$ і кількості ступенів вільності $k = r - l - 1 = 4 - 1 - 1 = 2$ знайдемо критичну точку правосторонньої критичної області (додаток 8):

$$\chi^2_{кр}(\alpha; k) = \chi^2_{кр}(0,05; 2) = 6.$$

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

12.3. Алгоритми перевірок гіпотез для дискретних випадкових величин

12.3.1. Перевірка гіпотези про розподіл генеральної сукупності за біноміальним законом

Для перевірки нульової гіпотези H_0 про біноміальний розподіл генеральної сукупності за рівня значущості α, виконуємо такі кроки:

1) висуваємо гіпотезу H_0: емпіричний розподіл відповідає біноміальному; альтернативна гіпотеза H_1: емпіричний розподіл не відповідає біноміальному;

2) рівень значущості приймаємо α (якщо він не заданий за умовою);

3) застосовуємо критерій χ^2 Пірсона, для чого обчислимо $\chi^2_{сп}$ за формулою: $\chi^2_{спост} = \sum_{i=1}^{r} \frac{\left(n_i - n_i'\right)}{n_i'}$, де $n_i' = np_i$, p_i знаходимо за формулою Бернуллі $p_i = P\left(X = i\right) = C_n^i p^i q^{n-i}$;

4) знайдемо $\chi^2_{кр}$ за таблицею критичних точок розподілу $\chi^2\left(\alpha; k\right)$; де $k = r - l - 1$, а r — кількість груп вибірки, що залишилася після об'єднання малочисельних груп, $l = 1$ або $l = 0$ залежно від того, оцінювався параметр $\left(p\right)$ за вибіркою, чи ні;

5) зіставляємо $\chi^2_{кр}$ і $\chi^2_{спост}$. Якщо $\chi^2_{кр} > \chi^2_{спост}$, то немає підстав, щоб відхилити нульову гіпотезу; якщо ж $\chi^2_{кр} \le \chi^2_{спост}$, то нульову гіпотезу відхиляємо.

Зауваження. Якщо емпіричні частоти малочисельні $\left(< 5\right)$, то їх і відповідні їм теоретичні частоти об'єднуємо з попередніми чи наступними значеннями. При цьому кількість груп вибірки зменшиться.

Приклад 12.4. У бібліотеці випадковим чином відібрано 300 комплектів по 6 книг. При цьому реєстрували кількість пошкоджених книг. У результаті одержали емпіричний розподіл, що його подано в таблиці:

x_i	0	1	2	3	4	5	6
n_i	85	93	65	29	16	9	3

За умовою задачі x_i — кількість пошкоджених книг в одному комплекті; n_i — частота, тобто кількість комплектів, які містять хі пошкоджених книг.

Використовуючи критерій Пірсона, для рівня значущості 0,005 перевірити гіпотезу про те, що дискретна випадкова величина X — кількість пошкоджених книг, розподілена за біноміальним законом.

Розв'язання:

1. Обчислимо частоту ω та візьмемо її за оцінку ймовірності того, що випадково відібрана книга виявиться пошкодженою. За формулою Бернуллі

$P_i = P(X = i) = C_n^i p^i q^{n-i}$ і знайдемо ймовірності $P_i \left(i = \overline{0,6}\right)$ того, що подія, яку розглядаємо в задачі, з'явиться в $n = 6$ випробуваннях рівно і раз. Для цього всі дані занесемо до таблиці. Маємо:

$$p = \omega = \frac{\sum_{i=0}^{6} \omega_i \cdot i}{6} = \frac{\frac{\sum_{i=0}^{6} n_i \cdot i}{n}}{6} = \frac{0 \cdot 85 + 1 \cdot 93 + 2 \cdot 65 + 3 \cdot 29 + 4 \cdot 16 + 5 \cdot 9 + 6 \cdot 3}{300 \cdot 6} \approx 0,24.$$

Оскільки $p \approx 0,24$; $q = 1 - p = 0,76$, тоді:

$p_0 = P_6(0) = C_6^2 p^0 q^6 = 0,1927$;

$p_1 = P_6(1) = C_6^1 p^1 q^5 = 6 \cdot 0,24 \cdot 0,76^5 \approx 0,3651$;

$p_2 = P_6(2) = C_6^2 p^2 q^4 = \frac{6 \cdot 5}{2} \cdot 0,24^2 \cdot 0,76^4 \approx 0,1927$;

$p_3 = \frac{6 \cdot 5 \cdot 4}{2 \cdot 3} \cdot 0,24^3 \cdot 0,76^3 \approx 0,1214$;

$p_4 = \frac{6 \cdot 5}{1 \cdot 2} \cdot 0,24^4 \cdot 0,76^2 \approx 0,0287$;

$p_4 = 6 \cdot 0,24^5 \cdot 0,76 \approx 0,1214$; $p_6 = 0,24^6 \approx 0,0002$.

i	n_i	$i \cdot n_i$	p_i	$n_i' = np_i = 300 \cdot p_i$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n'}$
0	85	0	0,1927	57,81	27,19	12,79
1	93	93	0,3651	109,53	273,24	2,495
2	65	130	0,2882	86,46	460,532	5,327
3	29	87	0,1214	36,42	55,056	1,512
4	16	64	0,0287	8,61	54,61	6,343
5 6	$\left.\begin{matrix}9\\3\end{matrix}\right\}12$	$\left.\begin{matrix}45\\18\end{matrix}\right\}63$	$\left.\begin{matrix}0,0037\\0,0002\end{matrix}\right\}0,0039$	$\left.\begin{matrix}1,11\\0,06\end{matrix}\right\}1,17$	117,29	100,25
$\sum$	300	200	1	300	-	$\chi^2_{спост} = 128,717$

3. Визначимо $\chi^2_{кр}$ за рівнем значущості $\alpha = 0,05$ і числом ступенів свободи $k = r - l - 1 = 6 - 1 - 1 = 4$, ($s = 6$ оскільки останні дві групи об'єднанні завдяки тому, що остання група малочисельна); тоді:

$\chi^2_{кр} = \chi^2(0,05;4) = 9,5$.

4. Оскільки $\chi^2_{кр} < \chi^2_{спост} \Rightarrow 9,5 < 128,717$, то гіпотеза H_0 про відповідність емпіричного розподілу біноміальному відхиляється.

12.3.2. Перевірка гіпотези про розподіл випадкової величини за законом Пуассона

У цьому разі діємо аналогічно до попереднього випадку, тільки теоретичні частоти обчислюємо за формулою:

$$n'_m = np_m = n \cdot P(X = m) = \frac{n \cdot \overline{x}_B^m}{m!} e^{-\overline{x}_B}$$

Для знаходження ймовірностей $P(X = m)$ може бути використана таблиця ймовірностей розподілу Пуассона. Якщо для обчисленого значення $\overline{x}_B$ ймовірності $P(X = m)$ нетабульовані, то для полегшення обчислень можна застосувати рекурентну формулу:

$$P(X = m+1) = \frac{\overline{x}_B}{m+1} P(X = m);\ P(X = 0) = e^{-\overline{x}_B}.$$

Приклад 12.5. Проведено спостереження за кількістю викликів на телефонній станції. З цією метою протягом 100 випадково вибраних 5-секундних інтервалів часу реєструвалася кількість викликів. Одержали такий варіаційний ряд:

Кількість викликів, x_i	0	1	2	3	4	5
Частоти, n_i	8	28	31	18	9	6

Використовуючи критерій Пірсона, перевірити гіпотезу про те, що розподіл кількості викликів χ узгоджується із законом Пуассона. Рівень значущості прийняти $\alpha = 0,05$.

Розв'язання: $n = \sum n_i = 100$.

Крок 1. Гіпотеза H_0: випадкова величина X розподілена за законом Пуассона.

Крок 2. Обчислимо ймовірність m викликів протягом n випадково вибраних відрізків часу за формулою Пуассона:

$$P_n(m)=\begin{cases}\dfrac{\lambda^m}{m!}e^{-\lambda}, & \text{якщо} \quad m=\overline{0,n}\\ 0, & \text{в усіх інших випадках.}\end{cases}$$

Крок 3. Знайдемо точкову оцінку параметра λ. Як відомо, точковою оцінкою для математичного сподівання генеральної сукупності є $\bar{x}_B$, обчислимо її:

$$\bar{x}_B=\frac{\sum_{i=1}^{6}x_i n_i}{n}=\frac{0\cdot 8+1\cdot 28+2\cdot 31+3\cdot 18+4\cdot 9+5\cdot 6}{100}=2{,}1$$

Тоді функція ймовірностей закону, який допускаємо у гіпотезі H_0, набуде вигляду:

$$P_n(m)=\frac{2{,}1^m\cdot e^{-2{,}1}}{m!}.$$

Крок 4. Застосовуємо критерій χ^2 Пірсона: $\chi^2=\sum_{i=1}^{r}\frac{(n_i-n_i')^2}{n_i'}$.

Знайдемо $\chi^2_{спост}$, для чого всі обчислення для зручності здійснимо в таблиці:

x_i	n_i	$p_i=P_n(m)$	$np_i=n_i'$	$(n_i-n_i')^2$	$\frac{(n_i-n_i')^2}{n_i'}$
0	8	0,1225	12,25	18,0625	1,4745
1	28	0,2572	25,72	5,1984	0,2021
2	31	0,2700	27	16	0,5926
3	18	0,1890	18,9	0,81	0,0429
4	9	0,0992	9,92	0,8464	0,0853
5	6	0,0621	6,21	0,0441	0,0071
$\sum$	100	1	100	-	$\chi^2_{спост}=2{,}4045$

Для останнього значення $p_5=1-\sum_{i=0}^{4}p_i=0{,}0621$

Крок 5. Із таблиці значень критичних точок розподілу χ^2 за даним рівнем значущості $\alpha = 0,05$ і числом ступенів вільності $k = r - l - 1 = 6 - 1 - 1 = 4$ знаходимо, що $\chi^2_{кр}(0,05; 4) = 9,5$.

Оскільки $\chi^2_{спост} = 2,4045 < \chi^2_{кр} = 9,5$, тому немає підстав для відхилення нульової гіпотези про те, що кількість викликів на телефонній станції розподілена за законом Пуассона.

Приклад 12.6. У бригаді 5 людей. Протягом 100 днів фіксували кількість робітників, що перевиконували план. Дістали такі результати:

Кількість i робітників, що перевиконали план	1	2	3	4	5
Кількість зафіксованих випадків, n_i	31	23	11	8	2

Перевірити за допомогою критерію Пірсона при рівні значущості $\alpha = 0,05$ гіпотезу про розподіл Пуассона кількості робітників, які перевиконували план.

Розв'язання. За розподілом Пуассона ймовірність того, що випадкова величина X набула значення, яке дорівнює i, обчислюється за формулою:

$P(i,a) = \dfrac{e^{-a}a^i}{i!}$, де a - математичне сподівання даної випадкової величини.

За даними спостереження обчислюємо статистичну оцінку $\overline{x}_B$ параметра a закону розподілу Пуассона за формулою:

$$\overline{x}_B = \frac{\sum_{i=1}^{5} i n_i}{n} = \frac{1\cdot 31 + 2\cdot 23 + 3\cdot 11 + 4\cdot 8 + 5\cdot 2}{31 + 23 + 11 + 8 + 2} = \frac{152}{75} = 2,0267;$$

$$e^{-2,0267} = 0,1318.$$

Оскільки розподіл Пуассона прямує від 0 до $+\infty$, то потрібно брати $i = 0$, відповідно $p_0 = e^{-a} = e^{-2,0267} \approx 0,1318$. Інші значення, окрім останнього, обчислюємо згідно формули: $P(i;2,0267) = \frac{0,1318 \cdot 2,0267^i}{i!}$. Для останнього значення $p_5 = 1 - \sum_{i=0}^{4} p_i = 0,0551$.

Подальші обчислення занесемо до таблиці.

i	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
0	0,1318	9,8850	-9,8850	97,7132	9,8850
1	0,2671	20,0325	10,9675	120,2861	6,0045
2	0,27076	20,295	2,705	7,3170	0,3605
3	0,1828	13,7100	-2,7100	7,3441	0,5357
4	0,0926	6,9450	1,0550	1,1130	0,1603
5	0,0551	4,1325	0,8675	0,7526	0,1821
	1	75			17,1281

$\chi^2 = 17,1281$.

Оскільки кількість ступенів свободи $k = r - l - 1$, де $r = 6$, $l = 1$ (лише параметр a), то $k = 4$. За таблицею розподілу χ^2 при $k = 4$ і $\alpha = 0,05$ знаходимо $\chi^2_\alpha = 9,5$ (дод. 8). Гіпотеза про те, що кількість робітників, які перевиконували план, розподілена за законом Пуассона, не може бути прийнятою, оскільки $\chi^2 > \chi^2_\alpha$.