Лекція 4 (продовження).

Перевірка статистичних гіпотез

4.6. Перевірка гіпотез про параметр p біноміального розподілу

При статистичному аналізі даних підпорядкованих схемі Бернуллі, розглядають два типи задач: порівняння ймовірності успіху p в одному випробуванні із заданим значенням p_0 і порівняння ймовірностей успіху у двох серіях випробувань.

1. В першому випадку перевіряємо гіпотезу $H_0 : p = p_0$. Нехай в n випробуваннях за схемою Бернуллі «успіх» відбувся μ разів. В якості статистики критерію вибирають відносну частоту $\omega = \frac{\mu}{n}$. При великих $n\left(n > 50\right)$ і при виконанні умови $n\omega > 5,\ n\left(1-\omega\right) > 5$ розподіл випадкової величини ω з достатньою для практичних розрахуунків точністю апроксимується розподілом $N\left(p; \sqrt{\frac{p\left(1-p\right)}{n}}\right)$. Тому при перевірці $H_0 : p = p_0$ статистика $z = \frac{\frac{\mu}{n} - p_0}{\sqrt{p_0\left(1-p_0\right)}}\sqrt{n}$ має розподіл $N\left(0;1\right)$.

Схема перевірки нульової гіпотези:

1. Нульова гіпотеза $H_0 : p = p_0$.

2. Альтернативна гіпотеза:

а) $H_1 : p \neq p_0$;

б) $H_1 : p = p_1 > p_0$;

в) $H_1 : p = p_1 < p_0$.

3. Критерій: $z = \dfrac{\dfrac{\mu}{n} - p_0}{\sqrt{p_0\left(1 - p_0\right)}}\sqrt{n}$.

4. Критичні точки:(залежать від α)

а) Двостороння критична область: критичні точки симетричні $z_{1кр} = -z_{2кр}$ і знаходяться з формули: $z_{\frac{\alpha}{2}} = \Phi^{-1}\left(0,5 - 0,025\right) = 1,96$ за таблицею Лапласа при $\alpha = 0,05$.

б) Правостороння критична область: $\Phi\left(z_{кр,\alpha}\right) = \dfrac{1 - 2\alpha}{2}$, $z_{0,05} = 1,645$.

в) Лівостороння критична область: $-z_{0,05} = -1,645$.

5. Правило прийняття рішень: гіпотеза H_0 відхиляється:

а) $\left|z_{спост}\right| \geq z_{кр,\frac{\alpha}{2}}$;

б) $z_{спост} \geq z_{кр,\alpha}$;

в) $z_{спост} \leq -z_{кр,\alpha}$.

Приклад 4.9. Директор великої страхової компанії цікавиться плинністю кадрів страхових агентів у перший рік роботи. Аналіз показав, що 25% страхових агентів серед зарахованих на роботу звільняється, не відпрацювавши року. Для зарахованих на роботу 150 страхових агентів був проведений інтенсивний тренінг. В кінці року із цих 150 агентів звільнилося 29. Чи можна вважати, що відсоток відсоток агентів, які пройшли тренінг і звільнилися, менше 25% на рівні значущості 0,01?

Розв'язання.

1. Нульова гіпотеза $H_0 : p < 0,25$.

2. Критерій: $z = \dfrac{\dfrac{\mu}{n} - p_0}{\sqrt{p_0\left(1 - p_0\right)}}\sqrt{n}$.

   $z_{спост} = \dfrac{0,19 - 0,25}{\sqrt{0,25\left(1 - 0,25\right)}} = -1,7$, критична область лівостороння.

3. Критичні точки:(залежать від α**);** $\varPhi\left(z_{кр,0,01}\right) = \dfrac{1 - 2\cdot 0,01}{2} = 0,49$;

   $z_{0,01} = -2,33$.

4. Оскільки $z_{спост} = -1,7 > z_{кр,0,01} = -2,33$, то немає підстав відхиляти нульову гіпотезу.►

2. Для перевірки гіпотези $H_0 : p_1 = p_2$ про рівність параметрів біноміально розподілених сукупностей проводиться дві серії випробувань. Нехай деяка подія А в серії n_1 випробувань відбулася μ_1 разів, а в серії n_2 випробувань відбулася μ_2 разів $\left(n_1, n_2 \geq 100\right)$. Перевіримо гіпотезу про рівність ймовірностей настання події А в обох серіях випробувань. В якості оцінок невідомих ймовірностей беремо відповідні відносні частоти: $p_1 = \omega_1 = \dfrac{\mu_1}{n_1}$ і $p_2 = \omega_2 = \dfrac{\mu_2}{n_2}$.

Зазначимо, що при значній кількості випробувань $\omega_1 \to N\left(p_1, \frac{p_1 q_1}{n_1}\right)$,

$\omega_2 \to N\left(p_2, \frac{p_2 q_2}{n_2}\right)$. Будемо вважати, що при достатньо великих обсягах вибірок, вирази $\frac{(\mu_1+\mu_2)n_1}{n}$, $\frac{(\mu_1+\mu_2)n_2}{n}$, $\frac{(\kappa_1+\kappa_2)n_1}{n}$, $\frac{(\kappa_1+\kappa_2)n_2}{n}$ набуватимуть значень більших за 5;

$\left(n = n_1 + n_2;\ \kappa_1 = n_1 - \mu_1;\ \kappa_2 = n_2 - \mu_2\right)$.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0 : p_1 = p_2$ проти альтернативної $H_1 : p_1 \neq p_2$ потрібно обчислити спостережуване значення за формулою:

$$z_{спост} = \frac{\frac{\mu_1}{n_1} - \frac{\mu_2}{n_2}}{\sqrt{\frac{\mu_1+\mu_2}{n_1+n_2}\left(1 - \frac{\mu_1+\mu_2}{n_1+n_2}\right)\left(\frac{1}{n_1} - \frac{1}{n_2}\right)}}$$

і за таблицею Лапласа знайти критичне значення за формулою

$\Phi\left(z_{кр,\alpha}\right) = \frac{1-\alpha}{2}$. Якщо $\left|z_{спост}\right| \geq z_{кр,\frac{\alpha}{2}}$ нульова гіпотеза відхиляється.

Правило 2. При альтернативній гіпотезі $H_1 : p_1 > p_2$ знаходять критичну точку правосторонньої критичної області із рівності $\Phi\left(z_{кр,\alpha}\right) = \frac{1-2\alpha}{2}$.

Гіпотеза H_0 відхиляється, якщо $z_{спост} \geq z_{кр,\alpha}$.

Правило 3. При альтернативній гіпотезі $H_1 : p_1 < p_2$ знаходять критичну точку лівосторонньої критичної області із рівності $\Phi\left(z_{кр,\alpha}\right) = \frac{1-2\alpha}{2}$.

Гіпотеза H_0 відхиляється, якщо $z_{спост} < -z_{кр,\alpha}$.

Приклад 4.10. Досліджувалися дві партії типових виробів. Одержали наступні статистичні розподіли:

№ партії	Кількість виробів		Сума
	браковані	не браковані	
1	8	92	100
2	13	287	300
сума	21	397	400

Чи можна стверджувати, що частка браку в обох партіях одна і та сама при рівні значущості 0,05?

Розв'язання.

1. Нульова гіпотеза $H_0 : p_1 = p_2$.
2. Альтернативна гіпотеза: $H_1 : p_1 \neq p_2$.
3. $$z_{спост} = \frac{\frac{\mu_1}{n_1} - \frac{\mu_2}{n_2}}{\sqrt{\frac{\mu_1+\mu_2}{n_1+n_2}\left(1-\frac{\mu_1+\mu_2}{n_1+n_2}\right)\left(\frac{1}{n_1}-\frac{1}{n_2}\right)}} =$$

$$= \frac{\frac{8}{100} - \frac{13}{300}}{\sqrt{\frac{8+13}{100+300}\left(1-\frac{8+13}{100+300}\right)\left(\frac{1}{100}-\frac{1}{300}\right)}} = 1{,}44$$

критична область двостороння.

4. Критичні точки:(залежать від α); $\Phi\left(z_{кр,0,01}\right)=\dfrac{1-0,05}{2}=0,475$;

$\left|z_{0,05}\right|=1,96$.

5. Оскільки $\left|z_{спост}\right|=1,44<z_{кр,0,05}=1,96$, то немає підстав відхиляти нульову гіпотезу.►

4.7. Перевірка правильності непараметричних статистичних гіпотез. Обчислення теоретичних частот

Всі перевірки параметричних статистичних гіпотез грунтувалися на припущені, що ознака генеральної сукупності має нормальний розподіл ймовірностей і що за іншого розподілу висновки щодо статистичних гіпотез можуть бути хибними. Підгрунтям для висновків про характер гіпотетичного розподілу можуть бути форми полігону або гістограми.

Емпіричними називають частоти, які спостерігаються при реалізації вибірки, а **теоретичними** – які обчислені за формулами.

Дискретні закони розподілу.

Теоретичні частоти обчислюються за формулою $n_i^{теор}=np_i$, де n – обсяг вибірки; p_i - ймовірність спостережуваного значення x_i, яка обчислюється за умови, що відповідна ознака генеральної сукупності має взятий за припущенням закон розподілу ймовірностей.

Біноміальний розподіл.

$p_i=C_n^i p^i q^{n-i}$; $n_i^{теор}=np_i=nC_n^i p^i q^{n-i}$.

Приклад 4.11. Відділ технічного контролю перевірив *n*=100 партій виробів по *m*=10 виробів у кожній партії і при цьому одержав такий емпіричний розподіл дискретної ВВ *X* – числа бракованих деталей:

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

де x_i - число бракованих виробів в партії; n_i - кількість партій, які мають таку кількість бракованих деталей. Припускаючи, що ВВ X має біноміальний розподіл, обчислити теоретичні частоти цієї вибірки.

Розв'язання. Знайдемо відносну частоту бракованих виробів і візьмемо її в якості оцінки $\hat{p}$ - ймовірності того, що навмання вибраний виріб буде бракований.

$$\hat{p} = \frac{\sum_{i=0}^{8} n_i x_i}{nm} = \frac{0\cdot 2 + 1\cdot 3 + 2\cdot 10 + 3\cdot 32 + 4\cdot 26 + 5\cdot 20 + 6\cdot 12 + 7\cdot 5}{100\cdot 10} = 0,4$$

Ймовірності за формулою Бернуллі:

$$P_{10}\left(0\right) = \hat{q}^{10} = 0,6^{10} = 0,00605;$$

$$P_{10}\left(1\right) = C_{10}^{1}\hat{p}\hat{q}^{9} = 10\cdot 0,4\cdot 0,6^{9} = 0,0403;$$

$$P_{10}\left(2\right) = C_{10}^{2}\hat{p}^{2}\hat{q}^{8} = 45\cdot 0,4^{2}\cdot 0,6^{8} = 0,121;$$

$$P_{10}\left(3\right) = C_{10}^{3}\hat{p}^{3}\hat{q}^{7} = 120\cdot 0,4^{3}\cdot 0,6^{7} = 0,215;$$

$$P_{10}\left(4\right) = C_{10}^{4}\hat{p}^{4}\hat{q}^{6} = 210\cdot 0,4^{4}\cdot 0,6^{6} = 0,2508;$$

$$P_{10}\left(5\right) = C_{10}^{5}\hat{p}^{5}\hat{q}^{5} = 0,2007;\ P_{10}\left(6\right) = 0,1115; P_{10}\left(7\right) = 0,0425.$$

Обчислюємо теоретичні частоти:

$$n_0^{теор} = np_0 = 100\cdot 0,00605 = 0,605 \approx 1;$$

$$n_1^{теор} = np_1 = 100 \cdot 0,0403 = 4,03 \approx 4;$$

$$n_2^{теор} = np_2 = 100 \cdot 0,121 = 12,1 \approx 12;$$

$$n_3^{теор} = np_3 = 100 \cdot 0,215 = 21,5 \approx 22;$$

$$n_4^{теор} = np_4 = 100 \cdot 0,2508 = 25,08 \approx 25;$$

$$n_5^{теор} = np_5 = 100 \cdot 0,2007 = 20,07 \approx 20;$$

$$n_6^{теор} = np_6 = 100 \cdot 0,1115 = 11,15 \approx 11;$$

$$n_7^{теор} = np_7 = 100 \cdot 0,0425 = 4,25 \approx 4.$$

У підсумку маємо:

Емпіричні частоти	2	3	10	22	26	20	12	5
Теоретичні частоти	1	4	12	22	25	20	11	4

Оскільки розбіжність між емпіричними та теоретичними частотами невелика, то можна вважати, що ця ВВ має біноміальний розподіл.

Розподіл Пуассона: $n_i^{теор} = np_i = n\dfrac{\lambda^i \cdot e^{-\lambda}}{i!}$, $\lambda = a$.

Оскільки $\overline{x}_B$ є незміщеною статистичною оцінкою для математичного сподівання a, то $n_i^{теор} = n\dfrac{\left(\overline{x}_B\right)^i \cdot e^{-\overline{x}_B}}{i!}$.

Приклад 4.12. За результатами вибірки, проведеної з генеральної сукупності, ознака якої за припущенням має пуассонівський розподіл, дістали такий статистичний розподіл:

x_i	1	3	5	7	9
n_i	6	10	16	5	3

Знайти теоретичні частоти.

Розв'язання.

Обсяг вибірки: $n = 40$.

Вибіркове середнє: $\overline{x}_B = 4,45$. Приймаємо $\lambda = 4,45$.

Обчислюємо ймовірності для розподілу Пуассона:

$$P_{40}(1) = \frac{\lambda}{1!} e^{-\lambda} = 4,45 \cdot e^{-2} \cdot e^{-2} \cdot e^{-0,45} = 4,45 \cdot 0,1353^2 \cdot 0,6376 = 0,052 \ .$$

Значення e^{-2} i $e^{-0,45}$ знаходимо за таблицею значень функції e^{-x} в додатках підручників або в інтернеті. Аналогічно,

$P_{40}(3) = 0,1714$; $P_{40}(5) = 0,1697$; $P_{40}(7) = 0,080$; $P_{40}(9) = 0,022$.

Теоретичні частоти:

$n_1^{теор} = nP_{40}(1) = 40 \cdot 0,052 \approx 2$; $n_2^{теор} = nP_{40}(3) \approx 7$;

$n_3^{теор} = nP_{40}(5) \approx 7$; $n_4^{теор} = nP_{40}(7) \approx 3$; $n_5^{теор} = nP_{40}(9) \approx 1$.

У підсумку маємо:

Емпіричні частоти	6	10	16	5	3
Теоретичні частоти	2	7	7	3	1

Оскільки розбіжність між теоретичними і емпіричними частотами велика, то це ставить під сумнів припущення про розподіл Пуассона для генеральної сукупності.

Неперервні закони розподілу.

Якщо ознака X генеоальної сукупності маєнеперервний розподіл ймовірностей, то теоретичні частоти обчислюються за формулою:

$n_i^{теор} = np_i$, де n – обсяг вибірки; p_i - ймовірність того, що випадкова величина X потрапить в і-ий частковий інтервал. Вона обчислюється за формулою того закону, який припускаємо за даними вибірки.

Рівномірний розподіл.

1) Оцінюємо параметри a,b - кінці інтервалу, в якому спостерігалися можливі значення X. Позначимо оцінки: $\hat{a},\hat{b}$.

2) Щільність ймовірності рівномірного позподілу: $\varphi(x)=\dfrac{1}{\hat{b}-\hat{a}}$.

3) Теоретичні частоти: $n_1^{теор} = np(x)(x_1-\hat{a}) = \dfrac{n}{\hat{b}-\hat{a}}(x_1-\hat{a})$.

$$n_2^{теор} = n_3^{теор} = n_4^{теор} = \ldots = n_{k-1}^{теор} = \frac{n}{\hat{b}-\hat{a}}(x_i - x_{i-1}).$$

$$n_k^{теор} = \frac{n}{\hat{b}-\hat{a}}(\hat{b} - x_{k-1}).$$

Приклад 4.13. Проведено *n*=200 випробувань, при цьому деяка подія *A* відбулася у різні моменти часу. Результати випробувань звели в емпіричний розподіл:

Інтервал, $(x_{i-1}-x_i)$	Частота, n_i	Інтервал, $(x_{i-1}-x_i)$	Частота, n_i
2-4	21	12-14	14
4-6	16	14-16	21
6-8	15	16-18	22
8-10	26	18-20	18
10-12	22	20-22	25

Обчислити теоретичні частоти, вважаючи, що час настання події *A* розподілений рівномірно.

Розв'язання.

Знайдемо оцінки параметрів a,b: для рівномірного розподілу

$$M(x)=\frac{a+b}{2}=\bar{x}_B;\ \ \sigma^2=\frac{(b-a)^2}{12};\ \ \sigma=\frac{b-a}{2\sqrt{3}}=\sigma_B.$$

Маємо систему :

$$\begin{cases} \dfrac{a+b}{2} = \overline{x}_B \\ \dfrac{b-a}{2\sqrt{3}} = \sigma_B \end{cases} \Leftrightarrow \begin{cases} \hat{a} = \overline{x}_B - \sqrt{3}\sigma_B \\ \hat{b} = \overline{x}_B + \sqrt{3}\sigma_B \end{cases}$$

. Для обчислень створимо інтервальний статистичний розподіл:

Середина інтервала, x_i'	3	5	7	9	11	13	15	17	19	21	Σ
Частота, n_i	21	16	15	26	22	14	21	22	18	25	200

Обчислюємо звичайним чином: $\overline{x}_B = 12{,}31;$ $\sigma_B = 5{,}81$. Тоді,

$$\begin{cases} \hat{a} = \overline{x}_B - \sqrt{3}\sigma_B = 12{,}31 - 1{,}73 \cdot 5{,}81 = 2{,}26 \\ \hat{b} = \overline{x}_B + \sqrt{3}\sigma_B = 12{,}31 + 1{,}73 \cdot 5{,}81 = 22{,}16 \end{cases}.$$

Обчислюємо щільність ймовірностей: $\varphi(x) = \dfrac{1}{\hat{b} - \hat{a}} = 0{,}05$.

Теоретичні частоти: $n_1^{теор} = np(x)(x_1 - \hat{a}) = 200 \cdot 0{,}05(4 - 2{,}26) = 17{,}4;$

$n_2^{теор} = 200 \cdot 0{,}05(x_2 - x_1) = 200 \cdot 0{,}05(5-3) = 20 = n_3^{теор} = \ldots = n_9^{теор};$

$n_{10}^{теор} = 200 \cdot 0{,}05(\hat{b} - 19) = 10(22{,}36 - 19) = 23{,}6.$

$\sum n_i^{теор} = 201.$

Оскільки розбіжності між теоретичними і емпіричними частотами незначні, то можна вважати, що ця ВВ розподілена рівномірно.

Нормальний розподіл.

Теоретичні частоти обчислюємо за формулою:

$$n_i^{теор}=\frac{n\cdot h}{\sigma_B}\varphi(x_i)=\frac{n\cdot h}{\sigma_B}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_i-\overline{x}_B)^2}{2\sigma_B^2}}$$, де де n – обсяг вибірки; h - довжина часткового інтервалу; $\overline{x}_B$ - вибіркова середня; σ_B - вибіркове середнє квадратичне відхилення; $\varphi(x)$ - щільність ймовірностей для загального нормального розподілу. Через функцію Лапласа:

$$n_i^{теор}=n\left(\Phi\left(\frac{x_{i+1}-\overline{x}_B}{\sigma_B}\right)-\Phi\left(\frac{x_i-\overline{x}_B}{\sigma_B}\right)\right).$$

Приклад 4.14. Дана таблиця про розподіл за розміром проданого магазином чоловічого взуття напротязі одгого дня:

Розмір взуття, x_i	37	38	39	40	41	42	43	44
Число проданих пар, n_i	1	4	14	37	35	20	8	3

Обчислити відповідний ряд теоретичних частот в припущені, що ВВ має нормальний розподіл.

Розв'язання.

При розв'язанні вважаємо, що задану вибірку взято в припущені, що ВВ є неперервною, а значення взяті як середини відповідних інтервалів $(36{,}5;37{,}5)$, $(37{,}5;38{,}5)$, …, $(43{,}5;44{,}5)$.

Знаходимо $\overline{x}_B=40{,}705;\ \sigma_B=1{,}322$.

Теоретичні частоти подамо у таблиці:

Інтервали, (α_i,β_i)	n_i	$\Phi\left(\frac{\beta_i-\overline{x}_B}{\sigma_B}\right)$	$\Phi\left(\frac{\alpha_i-\overline{x}_B}{\sigma_B}\right)$	$\left(\Phi\left(\frac{x_{i+1}-\overline{x}_B}{\sigma_B}\right)-\Phi\left(\frac{x_i-\overline{x}_B}{\sigma_B}\right)\right)n$
$(36{,}5;37{,}5)$	1	-0,4923	-0,4992	$0{,}0069\cdot122\approx1$
$(37{,}5;38{,}5)$	4	-0,4523	-0,4923	$0{,}0400\cdot122\approx5$
$(38{,}5;39{,}5)$	14	-0,3188	-0,4523	$0{,}1335\cdot122\approx16$

$(39{,}5; 40{,}5)$	37	-0,0616	-0,3188	$0{,}2572 \cdot 122 \approx 31$
$(40{,}5; 41{,}5)$	35	0,2261	-0,0616	$0{,}2877 \cdot 122 \approx 35$
$(41{,}5; 42{,}5)$	20	0,4128	0,2261	$0{,}1867 \cdot 122 \approx 23$
$(42{,}5; 43{,}5)$	8	0,4828	0,4128	$0{,}0700 \cdot 122 \approx 9$
$(43{,}5; 44{,}5)$	3	0,4975	0,4828	$0{,}0152 \cdot 122 \approx 2$
Σ	122			122

Оскільки розбіжність між емпіричними та теоретичними частотами невелика, то можна вважати, що ця ВВ має нормальний розподіл.►

Показниковий розподіл.

Нехай задано емпіричний розподіл неперервної випадкової величини X у вигляді послідовності інтервалів $(x_i; x_{i+1})$ і відповідних їм частот n_i.

Теоретичні частоти: $n_i^{теор} = nP_i$, де $P_i = P(x_i < x < x_{i+1}) = e^{-\lambda x_i} - e^{-\lambda x_{i+1}}$.

$n_i^{теор} = n\left(e^{-\lambda x_i} - e^{-\lambda x_{i+1}}\right)$. Числове значення параметра $\lambda = \dfrac{1}{M(X)}$.

Оскільки $\overline{x}_B$ є незміщеною оцінкою *M(X)*, то $\lambda = \dfrac{1}{\overline{x}_B}$. Тоді теоретичні

моменти обчислюємо за формулою $n_i^{теор} = n\left(e^{-\frac{x_i}{\overline{x}_B}} - e^{-\frac{x_{i+1}}{\overline{x}_B}}\right)$.

Приклад 4.15. Заданий інтервальний статистичний розподіл вибірки:

Інтервал	0-6	6-12	12-18	18-24	24-30	30-36
n_i	45	24	13	9	7	2

Обчислити відповідний ряд теоретичних частот в припущені, що ВВ має показниковий розподіл.

Розв'язання.

Будуємо дискретний розподіл:

Середини інтервалів, x_i'	3	9	15	21	27	33
n_i	45	24	13	9	7	2

$\overline{x}_B = 9{,}9;\ \hat{\lambda} = 0{,}101$. Для обчистень теоретичних частот використовуємо таблицю значень функції e^{-x} .

x_i	x_{i+1}	n_i	$e^{-\hat{\lambda}x_i}$	$e^{-\hat{\lambda}x_{i+1}}$	$n_i^{теор} = n\left(e^{-\hat{\lambda}x_i} - e^{-\hat{\lambda}x_{i+1}}\right)$
0	6	45	1	0,545	45
6	12	24	0,545	0,2982	25
12	18	13	0,2982	0,1620	14
18	24	9	0,1620	0,0888	7
24	30	7	0,0888	0,0483	4
30	36	2	0,0483	0,0262	2
36	∞	-	0,0262	0	3

Оскільки розбіжність між емпіричними та теоретичними частотами невелика, то гіпотетично можна вважати, що ця ВВ має показниковий розподіл.►

4.8. Критерій узгодження Пірсона $\left(\chi^2\right)$.

Критерії, за якими перевіряють,узгоджуються чи ні статистичні дані з розглядуваними гіпотезами, називають **критеріями узгодження**. Статистикою критерія Пірсона обирається величина

$$\chi^2 = \sum_{i=1}^{к} \frac{\left(n_i - np_i\right)^2}{np_i} .$$

При заданому рівні значущості α перевірка нульової гіпотези здійснюється за планом:

1) Обчислюють теоретичні частоти $n_к^{теор}$.

2) Обчислюють спостережене значення $\chi^2_{спост} = \sum_{i=1}^{к} \frac{\left(n_i - np_i\right)^2}{np_i}$.

3) Обчислюють число ступенів свободи: $к = r - m - 1$, де r – число часткових інтервалів інтервального статистичного розподілу вибірки; т –

число параметрів, яким визначається закон розподілу ймовірностей генеральної сукупності згідно з нульовою гіпотезою.

4) За таблицею χ^2 - критерію, знаходять критичне значення $\chi^2_{кр}$ яке відповідає заданому рівню значущості і числу степенів свободи.

5) Якщо $\chi^2_{спост} < \chi^2_{кр}$, то немає підстав відхиляти гіпотезу H_0. В іншому вирадку, гіпотеза відкидається.

Приклад 4.16. За спостереженнями, наведеними у таблиці, за χ^2 - критерієм перевірити гіпотезу про те, що ВВ Х має нормальний розподіл $(\alpha = 0{,}05,\ n = 300)$.

x_i	3	5	7	9	11	13	15	17	19
n_i	23	39	38	46	37	35	34	39	18

Розв'язання. 1) $\bar{x}_B = 10{,}56;\ \sigma_B = 4{,}66$.

2) Теоретичні частоти:

x_i	n_i	$n_i^{теор} = \left(\Phi\left(\frac{x_{i+1} - \bar{x}_B}{\sigma_B}\right) - \Phi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right)\right) n$
3	23	13,82
5	39	25,30
6	38	38,49
9	46	48,65
11	37	51,16
13	35	44,87
15	33	32,79
17	30	19,0
19	18	9,98

3) $\chi^2_{спост} = \sum_{i=1}^{\kappa} \frac{\left(n_i - n_i^{теор}\right)^2}{n_i^{теор}} = 32{,}76$.

4) Для $\alpha = 0{,}05$, $n = 300$ і $\kappa = r - m - 1 = 9 - 2 - 1 = 6$ за таблицею

знаходимо $\chi^2_{кр} = 12{,}6$.

Висновки: $\chi^2_{спост} = 32{,}76 > \chi^2_{кр} = 12{,}6$, отже, еіпотеза про нормальний

розподіл відхиляється.

4.9. Непараметричні критерії

На практиці доводиться мати справу іх випадковими величинами , закон розподілу яких невідомий. В цьому випадку використовують непараметричні критерії, при застосуванні яких не потріюно робити ніяких припущень відносно закону росподілу досліджуваної величини. Слід зауважити, що непараметричні закони є менш ефективними.
Розглянемо деякі з них.

Критерій знаків.

Критерій знаків перевіряє гіпотезу про те, що дві вибірки $\left(x_1, \ldots, x_n\right)$ і $\left(y_1, \ldots, y_n\right)$однакового обсягу взято з однієї і тієї ж генеральної сукупності. Досліджують знаки різниці спарених результатів і обчислюють число тих знаків, яких менше. Позначимо їх число через $\boldsymbol{r}$. Якщо нульова гіпотеза справедлива, то $P\left(x - y > 0\right) = P\left(x - y < 0\right) = \frac{1}{2}$.

Критерій Вілкоксона, Манна, Уітні.

Застосовують критерій для порівняння нгзалежних виюірок обсягами п1 і п2 і перевіряють гіпотезу, яка стверджує, що вибірки одержані із однорідних генеральних сукупностей і мають рівні середні і моди.

Контрольні питання

1. В чому полягає загальна схема перевірки статистичної гіпотези?
2. Для чого використовують положення теореми Неймана-Пірсона?
3. Чи можна провести перевірку гіпотези про рівність середніх двох сукупностей в припущені, що дисперсії невідомі?
4. Де зустрічаються гіпотези про числові значення параметрів?
5. Для чого бажано проводити побудову теоретичного закону розподілу за результатами дослідних даних?