МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ
КРЕМЕНЧУЦЬКИЙ ЛЬОТНИЙ КОЛЕДЖ

Циклова комісія природничих дисциплін

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни «Теорія ймовірностей і математична статистика»
обов’язкових компонент освітньо-професійної програми
першого (бакалаврського) рівня вищої освіти
Облік і аудит

за темою №7− Перевірка статистичних гіпотез

Харків 2022

ЗАТВЕРДЖЕНО
Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 30.08.2022 №8

СХВАЛЕНО
Методичною радою Кременчуцького
льотного коледжу Харківського
національного університету
внутрішніх справ
Протокол від 22.08.2022 №1

ПОГОДЖЕНО
Секцією Науково-методичної ради
ХНУВС з технічних дисциплін
Протокол від 29.08.2022 №8

Розглянуто на засіданні циклової комісії природничих дисциплін, протокол
від 10.08.2022 № 1

Розробник: доцент циклової комісії природничих дисциплін, к.ф. –м.н.
Семенов В.О.

Рецензенти:
1. Доцент кафедри автомобілі та трактори Кременчуцького національного університету імені Михайла Остроградського, кандидат технічних наук Черниш А.А.
2. Ст.викладач циклової комісії економіки та управління Кременчуцького льотного коледжу Харківського національного університету внутрішніх справ, кандидат економічних наук Цимбалістова О.А.

План лекції

1. Статистичні гіпотези. Основний принцип перевірки статистичних гіпотез.
2. Перевірка гіпотез про чисельні значення параметрів.
3. Критерій узгодження Пірсона (нормальний розподіл).

Рекомендована література:

Основна

1. Барковський В. В. Теорія ймовірностей та математична статистика ; 5-те видання : навчальний посібник / В. В. Барковський, Н. В. Барковська, О. К. Лопатін. – К. : Центр учбової літератури, 2010. – 424 с.
2. Волощенко А. Б. Теорія ймовірностей і математична статистика : навчаль-но-методичний посібник для самостійного вивчення дисципліни / А. Б. Волощенко, І. А. Джалладова. – К. : КНЕУ, 2003. – 256 с.
3. Жильцов О.Б. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с.
4. Конспект лекцій.

Інформаційні ресурси в інтернеті

5. Жильцов О.Б. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с. [https://yandex.ua/search/?lr=143&oprnd=9346860898&text=жильцов%20теорія%20ймовірностей

Текст лекції

1. Статистичні гіпотези. Основний принцип перевірки статистичних гіпотез.

Статистичною називають гіпотезу про вигляд невідомого розподілу або про невідомі параметри відомих розподілів.

Нульовою (основною) гіпотезою H_0 називають висунуту гіпотезу. *Альтернативною (конкуруючою) гіпотезою* H_1 називають протилежну до нульової гіпотезу.

Похибка першого роду полягає у відхиленні правильної нульової гіпотези внаслідок її перевірки. *Рівень значущості* α — це ймовірність похибки першого роду (число $\alpha > 0$, проте близьке до нуля). *Похибка другого роду* полягає у прийнятті неправильної нульової гіпотези внаслідок її перевірки. Імовірність похибки другого роду позначають β.

Статистичним критерієм (або просто критерієм) називають випадкову величину K, яка використовується для перевірки нульової гіпотези. *Спостережуваним (емпіричним)* значенням $K_{сп}$ є значення критерію, обчислене за вибірками.

Критична область — це сукупність значень критерію, при яких відхиляють нульову гіпотезу. *Областю прийняття гіпотези (областю допустимих значень)* є сукупність значень критерію, при яких приймають нульову гіпотезу.

Основний принцип перевірки статистичних гіпотез такий: якщо спостережуване значення критерію належить критичній області, нульову гіпотезу відхиляють; якщо спостережуване значення критерію належить області прийняття гіпотези, гіпотезу приймають.

Критичними точками (межами) $k_{кр}$ називають такі, що відділяють критичну область від області прийняття гіпотези. *Правостороння критична область* визначається нерівністю $K > k_{кр}$. *Лівостороння критична область* визначає-ться нерівністю $K < k_{кр}$. *Двостороння критична область* визначається нерів-ностями $K < k_1$, $K > k_2$, де $k_1 < k_2$. Зокрема, якщо критичні точки симетричні відносно нуля, то двостороння область визначається нерівностями $K < -k_{кр}$, $K > k_{кр}$ (вважається, що $k_{кр} > 0$) або рівносильною нерівністю $|K| > k_{кр}$. Для відшукання критичної області задають рівень значущості і знаходять критичні точки із таких співвідношень:

а) для правосторонньої критичної області $P(K > k_{кр}) = \alpha$, $k_{кр} > 0$;

б) для лівосторонньої критичної області $P(K < k_{кр}) = \alpha$, $k_{кр} < 0$;

в) для двосторонньої симетричної критичної області

$$P(K < -k_{кр}) = \frac{\alpha}{2};\quad P(K > k_{кр}) = \frac{\alpha}{2};\quad k_{кр} > 0.$$

Потужністю критерію називають ймовірність того, що значення критерію попаде в критичну область за умови, що правильною є конкуруюча гіпотеза. Іншими словами, потужність критерію — це ймовірність того, що нульова гіпотеза буде відхилена, якщо справджується конкуруюча гіпотеза.

2. Перевірка гіпотез про чисельні значення параметрів.

2.1 Перевірка гіпотези про значення генерального середнього нормальної генеральної сукупності.

Нехай дано нормальну генеральну сукупність (тобто множину значень нормально розподіленої випадкової величини X) з відомою дисперсією σ^2 і невідомим генеральним середнім (математичним сподіванням) a. З цієї сукупності отримано вибірку об'єму n і підраховано вибіркове середнє $\overline{x}_в$.

Правило. Для того, щоб при заданому рівні значущості α перевірити нульову (або основну) гіпотезу $H_0 : a = a_0$ при конкуруючій гіпотезі $H_1 : a \neq a_0$, потрібно обчислити спостережуване значення критерію $Uсп = \frac{(\overline{x}_в - a_0)\sqrt{n}}{\sigma}$ і за таблицею значень функції Лапласа знайти критичну точку $u_{кр}$ двосторон-ньої критичної області з рівності $\Phi(u_{кр}) = \frac{1-\alpha}{2}$.

У випадку | $U_{сп}$ | ≤ $u_{кр}$ немає підстав відхиляти нульову гіпотезу, а у випадку | $U_{сп}$ | > $u_{кр}$ нульова гіпотеза не узгоджується із статистичними даними, і тому приймається конкуруюча гіпотеза.

***Приклад* 1.** Із нормальної генеральної сукупності з відомим середнім квадра-тичним відхиленням σ = 4 отримано вибірку об'єму n = 100, за якою знай-дено вибіркове середнє $\overline{x}_в$ = 29 . Потрібно при рівні значущості 0,01 переві-рити нульову гіпотезу H_0 : $a = a_0 = 30$ при конкуруючій гіпотезі H_1 : $a \neq 30$.

Розв'язання. Знайдемо спостерігаєме значення критерію

$$Uсп = \frac{(29-30)\sqrt{100}}{4} = -2,5.$$

Оскільки за умовою конкуруюча гіпотеза має вигляд H_1 : $a \neq a_0$, критична область -двостороння. Обчислимо критичну точку з рівності $\varPhi(u_{кр}) = \frac{1-0,01}{2} = 0,495$, а за таблицею значень функції Лапласа знаходимо критичну точку: $u_{кр}$ ≈ 2,58. Оскільки | $U_{сп}$ | < $u_{кр}$, підстав відхиляти нульову гіпотезу немає. Іншими словами, нема підстав вважати, що вибіркове та гіпотетичне генеральні середні різняться суттєво.

2.2. Перевірка гіпотези про значення дисперсії нормальної генеральної сукупності

Нехай з нормальної генеральної сукупності отримано вибірку об'єму n, для якої знайдено «виправлену» вибіркову дисперсію s^2.

Правило. Для того щоб при заданому рівні значущості α перевірити нульову гіпотезу H_0 : $\sigma^2 = \sigma_0^2$ про рівність невідомої генеральної дисперсії σ^2 гіпотетичному (пропонованому) σ_0^2 при конкуруючій гіпотезі H_1 : $\sigma^2 > \sigma_0^2$ потрібно обчислити спостережуване значення критерію $\chi^2_{сп} = \frac{(n-1)s^2}{\sigma_0^2}$ і за таблицею критичних точок розподілу Пірсона χ^2 , при заданому рівні значущості α і кількості ступенів свободи $k = n — 1$ знайти критичну точку $\chi^2_{кр}(\alpha, k)$. Якщо $\chi^2_{сп} < \chi^2_{кр}$, немає підстав відхиляти нульову гіпотезу. Якщо $\chi^2_{сп} > \chi^2_{кр}$, нульову гіпотезу відхиляють.

***Приклад* 2.** Із нормальної генеральної сукупності отримано вибірку об'єму n = 20, для якої знайдено вибіркову «виправлену» дисперсію s^2 = 2,7. Перевірити при рівні значущості α = 0,05 гіпотезу H_0 : $\sigma^2 = \sigma_0^2 = 3$, якщо альтернативна гіпотеза H_1 : $\sigma^2 > \sigma_0^2$

Розв'язання. Знайдемо спостережуване значення критерію:

$$\chi^2_{сп} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(20-1)\cdot 2,7}{3} = 17,1.$$

За таблицею значень критерію Пірсона χ^2 при рівні значущості $\alpha = 0,05$ і кількості степенів свободи $k = n - 1 = 20 - 1 = 19$ знайдемо критичну точку: $\chi^2_{кр}(0,05;19) = 30,14$.Оскільки $\chi^2_{сп} < \chi^2_{кр}$, немає підстав відхиляти основну гіпотезу про рівність генеральної дисперсії гіпотетичному значенню $\sigma_0^2 = 3$. Іншими словами, є підстави вважати, що різниця між «виправленою» вибірковою дисперсією $s^2 = 2,7$ і гіпотетичною генеральною дисперсією $\sigma_0^2 = 3$ незначуща.

3. Критерій узгодження Пірсона (нормальний розподіл)

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

x_i	x_1	x_2	...	x_N
n_i	n_1	n_2	...	n_N

Необхідно за допомогою *критерію Пірсона* перевірити гіпотезу про нормальний розподіл генеральної сукупності.

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

1) обчислити вибіркове середнє $\overline{x}_в$ і вибіркове середнє квадратичне відхилен-ня $\sigma_в$;

2) визначити теоретичні частоти $n'_i = \dfrac{nh}{\sigma_в} \cdot \varphi(u_i)$, де n — об'єм вибірки; h — крок (різниця між двома сусідніми варіантами); $\varphi(u)$ — значення диференціальної функції Лапласа для $u_i = \dfrac{x_i - \overline{x}_в}{\sigma_в}$;

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

а) складають розрахункову таблицю, за якою знаходять спостережуване зна-чення критерію Пірсона $\chi^2_{сп} = \sum_i \dfrac{(n_i - n'_i)^2}{n'_i}$;

б) за таблицею критичних точок розподілу Пірсона χ^2 при заданому рівні значущості α і кількості степенів свободи $k = s-3$ (s - кількість варіант ви-бірки) знаходять критичну точку $\chi^2_{кр}(\alpha, k)$ правосторонньої критичної області.

Якщо $\chi^2_{сп} < \chi^2_{кр}$, немає підстав відхиляти гіпотезу про нормальний розподіл генеральної сукупності, тобто вважати, що емпіричні та теоретичні частоти різняться суттєво. Якщо $\chi^2_{сп} > \chi^2_{кр}$, гіпотезу відхиляють, тобто вважають, що емпіричні та теоретичні частоти різняться суттєво.

Зауваження. Малочисельні частоти вибірки ($n_i < 5$) слід об'єднати; у цьому разі відповідні їм теоретичні частоти також потрібно додати. Якщо відбувалося об'єднання частот, то при визначенні кількості степенів свободи за формулою $k = s - 3$ слід за s узяти кількість варіант вибірки, що залиши-лася після об'єднання частот.

Приклад 3. Використовуючи критерій Пірсона, при рівні значущості 0,05 перевірити, чи справджується гіпотеза про нормальний розподіл генеральної сукупності X, якщо з цієї сукупності отримано вибірку об'єму $n = 100$:

x_i	-5	-3	-1	1	3	5	7	9
n_i	5	9	13	18	21	18	10	6

Розв'язання. Обчислимо вибіркове середнє $x_в$ і середнє квадратичне відхилення $\sigma_в$:

$$\overline{x}_{\text{в}} = \frac{-5\cdot 5 - 3\cdot 9 - 1\cdot 13 + 1\cdot 18 + 3\cdot 21 + 5\cdot 18 + 7\cdot 10 + 9\cdot 6}{100} = 2,3;$$

$$\sigma_{\text{в}} = \left(\left[5^2\cdot 5 + 3^2\cdot 9 + 1\cdot 13 + 1\cdot 18 + 3^2\cdot 21 + 5^2\cdot 18 + 7^2\cdot 10 + 9^2\cdot 6\right]/100 - \right.$$

$$\left.-2,3^2\right)^{\frac{1}{2}} = 3,64.$$

Визначимо теоретичні частоти, ураховуючи, що $n = 100$, $h = 2$, $\sigma_в = 3,64$, за формулою

$$n_i' = \frac{nh}{\sigma_в}\cdot\varphi\left(\frac{x_i - \overline{x}_в}{\sigma_в}\right).$$

Для зручності складемо розрахункову таблицю (значення функції Лапласа $\varphi(u)$ беремо з відповідної таблиці у додатку).

i	x_i	$u_i = \dfrac{x_i - \overline{x}_{\text{в}}}{\sigma_{\text{в}}}$	$\varphi(u_i)$	$n_i' = \dfrac{nh}{\sigma_{\text{в}}}\cdot\varphi(u_i)$
1	–5	–2,01	0,0529	2,91
2	–3	–1,46	0,1374	7,56
3	–1	–0,91	0,2637	14,50
4	1	–0,36	0,3739	20,56
5	3	0,19	0,3918	21,54
6	5	0,74	0,3034	16,68
7	7	1,29	0,1736	9,55
8	9	1,84	0,0734	4,04

Необхідно порівняти емпіричні та теоретичні частоти. Для цього обчислимо спостережуване значення критерію $\chi^2_{сп} = \sum_i \frac{(n_i - n_i')^2}{n_i'}$; Обчислення зручно оформити у вигляді 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	5	2,91	2,09	4,37	1,50
2	9	7,56	1,44	2,07	0,27
3	13	14,50	–1,50	2,25	0,16
4	18	20,56	–2,56	6,55	0,32
5	21	21,54	–0,54	0,29	0,01
6	18	16.68	1,32	1,74	0,10
7	10	9,55	0,45	0,20	0,02
8	6	4,04	1,96	8,34	0,95
Σ	100				$\chi^2_{сп} = 3,33$

За таблицею критичних точок розподілу Пірсона χ^2 при рівні значущості α = 0,05 і кількості степенів свободи $k = s - 3 = 8 - 3 = 5$ знаходимо критичну точку правосторонньої критичної області: $\chi^2_{кр}(0,05;5) = 11,07$. Оскільки $\chi^2_{сп} < \chi^2_{кр}$, гіпотезу про нормальний розподіл приймаємо, тобто вважаємо, що емпіричні та теоретичні частоти різняться несуттєво.

За допомогою критерія Пірсона можна також перевірити гіпотези про рівно-мірний, показниковий та біноміальний розподіли генеральної сукупності.