<u>Одновибірковий критерій погодженості Колмогорова</u>

Нехай $x_1, \ldots, x_n$ - вибірка з генеральної сукупності ξ. Потрібно перевірити гіпотезу про те, що випадкова змінна ξ керується неперервною функцією розподілу $F(x)$:

$$H_0: \quad P\{\xi \le x\} = F(x).$$

Методика перевірки цієї нульової гіпотези грунтується на висновку теореми Глівенка, згідно якої:

$$P\left(\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \underset{n\to\infty}{\rightarrow} 0\right) = 1, \text{ де}$$

$F_n(x)$ - емпірична функція розподілу.

На основі цієї вибірки знаходимо емпіричну функцію розподілу:

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \dfrac{m_i}{n}, & x_{(i)} \le x < x_{(i+1)}, \quad i = 1,\ldots n-1 \\ 1, & x_{(n)} \le x, \end{cases}$$

де m_i - кількість елементів $x_{(k)}$ варіаційного ряду деякої вибірки, які не більші за x. Розглянемо статистику:

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F(x)|.$$

Колмогоров довів, що статистика $K_n = \sqrt{n} D_n$ має розподіл незалежний від гіпотетичної функції розподілу $F(x)$, який при $n \to \infty$ збігається до розподілу:

$$K(x) = \begin{cases} 0, & x < 0 \\ \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2x^2k^2}, & x \ge 0 \end{cases}$$

Статистику $K_n = \sqrt{n} D_n$ називають статистикою Колмогорова.

<u>Алгоритм критерію Колмогорова</u>

1) Вибираємо рівень значущості α.

2) Будуємо варіаційний ряд даної вибірки.

3) Знаходимо значення гіпотетичної функції розподілу в точках варіаційного ряду та модулі різниць значень емпіричної функції у точках безмежно близьких зліва до точок цього ряду та гіпотетичної функції розподілу в точках варіаційного ряду.

4) Знаходимо емпіричне значення $K_{емп}$ статистики Колмогорова як максимальне значення модулів таких різниць $D_n = \max(d^+, d^-)$,

де $d^+ = \max|F_n(x_i) - F(x_i)|$, $d^- = \max|F_n(x_i - 0) - F(x_i)|$).

5) При вибраному рівні значущості α, знаходимо критичне значення $K_{кр}$ статистики Колмогорова

α	0,1	0,05	0,01
$K_{кр}$	1,23	1,36	1,63

Критичні значення статистики Колмогорова для найбільш уживаних рівнів значущості.

6) Якщо $K_{емп} < K_{кр}$, то гіпотезу приймаємо, в протилежному випадку – відхиляємо, як хибну.

<u>Двовибірковий критерій погодженості Смірнова</u>

Нехай $x_1, \ldots, x_n$ - вибірка з випадкової змінної ξ, а незалежна від неї вибірка $y_1, \ldots, y_m$ - з випадкової змінної η. Потрібно перевірити таку нульову гіпотезу: обидві генеральні сукупності керовані однією і тією ж неперервною функцією розподілу, тобто ξ і η - стохастично еквівалентні.

На основі вибірок знаходимо емпіричні функції розподілу цих статистичних змінних:

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \dfrac{n_i}{n}, & x_{(i)} \le x < x_{(i+1)}, \quad i = 1, \ldots n-1 \\ 1, & x_{(n)} \le x, \end{cases}$$

де n_i - кількість елементів $x_{(i)}$ - кількість елементів варіаційного ряду першої вибірки, які не більші за *x* та:

$$G_m(x) = \begin{cases} 0, & x < y_{(1)} \\ \dfrac{m_i}{m}, & y_{(i)} \le x < y_{(i+1)}, \quad i = 1, \ldots m-1 \\ 1, & y_{(n)} \le x, \end{cases}$$

де n_i - кількість елементів $x_{(i)}$ - кількість елементів варіаційного ряду другої вибірки, які не більші за *x*.

Розглянемо статистику:

$$D_{nm} = \sup_{-\infty<x<\infty} |F_n(x) - G_m(x)|.$$

У 1999р. Смірнов довів, що статистика

$$S_{nm} = \sqrt{\frac{mn}{m+n}} D_{nm}.$$

при $n \to \infty, m \to \infty$ має розподіл Колмогорова.

<u>Алгоритм критерію Смірнова:</u>

1) Вибираємо рівень значущості α.
2) Будуємо спільний варіаційний ряд даних двох вибірок
3) Знаходимо $(S_{nm})_{емп}$.
4) Знаходимо $(S_{nm})_{кр} = K_{кр}$.
5) Якщо $(S_{nm})_{емп} < K_{кр}$, то висунуту гіпотезу приймаємо, в протилежному випадку – відхиляємо, як хибну.

Завдання для самостійної роботи

1. Дано вибірку незалежних спостережень над неперервною популяцією:
   x: 13,95; 14,18; 14,24; 14,25; 14,28; 14,33; 14,12; 14,06; 14,35; 14,00.

Перевірити гіпотезу про те, що дана вибірка взята з рівномірно розподіленої на проміжку [13,94; 14,44] генеральної сукупності.

2. Дано вибірку незалежних спостережень над неперервною змінною x:
   8,95; 9,24; 9,18; 9,28; 9,33; 9,25; 9,06; 9,12; 9,35; 9.

Перевірити гіпотезу про те, що дана вибірка взята з рівномірно розподіленої на проміжку [8,9; 9,4] генеральної сукупності.

3. Шийки робочої частини свердел шліфуються на шліфувальному верстаті. Номінал діаметра шийки повинен становити 9,8мм з технічним допуском 0,05мм. Вимірювання робочої частини шийки 20 свердел дали такі результати (мм):
   9,76; 9,77; 9,83; 9,79; 9,80; 9,81; 9,79; 9,75; 9,80; 9,78; 9,75; 9,81; 9,82; 9,76; 9,78; 9,77; 9,81; 9,79; 9,79; 9,77.
   Чи можна вважати, що номінал діаметра шийки дорівнює 9,8 мм з технічним допуском 0,05мм?
4. Дано вибірку 25 незалежних спостережень x над абсолютно неперервною випадковою змінною :

-0,65	-0,02	-0,30	0,37	-0,01	-1,54	-1,53	-0,26	0,44	0,21	0,27	0,28	1,00
0,60	1,25	0,88	1,43	-0,06	-0,81	-1,02	-0,86	0,77	-0,17	1,68	-1,17	

Перевірити гіпотезу про те, що генеральна сукупність, з якої взята дана вибірка, нормально розподілена з параметрами a=0, σ=1.

5. Маємо вибірку вимірювання діаметрів 20 втулок.

72,50 72,43 72,46 72,32 72,50 72,56 72,56 72,32 72,43 72,38 72,48 72,48 72,34 72,54 72,56 72,56 72,69 72,46 72,43 72,41

Перевірити, що статистична змінна рівномірно розподілена на проміжку [72,73].

6. Дано дві вибірки незалежних спостережень, над двома неперервними популяціями:

   ξ: 0,95 1,15 1,24 1,25 1,33 1,16 1,31 1,15 1,27 1,00

   η: 0,90 1,01 1,24 1,27 1,34 1,07 1,27 1,33 0,98 1,11

Використовуючи критерій Смирнова, перевірити гіпотезу про те, що популяції, з яких взяті вибірки, однаково розподілені.

7. Маємо результати вимірювань середньомісячної зимової температури в двох містах протягом 13 років:

A: -19,2 -14,8 -19,6 -11,1 -9,4 -16,9 -13,7 -4,9 -13,9 -9,4 -8,3 -7,9 -5,3

B: -21,8 -15,4 -20,8 -11,3 -11,6 -19,2 -13,0 -7,4 -15,1 -14,4 -11,1 -10,5 -7,2

Перевірити гіпотезу, що на розподіл температури впливають однакові (подібні) кліматичні умови.

8. Дано дві вибірки незалежних спостережень, над двома неперервними популяціями:

   ξ: 2,95 1,25 2,65 1,74 2,00 1,95 1,65 2,15 1,25 1,85

   η: 1,15 2,00 1,29 1,95 2,05 2,74 1,85 1,90 2,65 2,74

Використовуючи критерій Смірнова, перевірити гіпотезу про те, що популяції, з яких взяті вибірки, однаково розподілені.