

О.Г.Гнатишин, Б.В.Москвяк
КРИТЕРІЙ СМІРНОВА ДЛЯ ПОРІВНЯННЯ
ДВОХ ЗРІЗАНИХ ВИБІРОК

Нехай

$$x(\bar{1}, m) \leq \dots \leq x(\bar{l}, m) \leq \dots \leq x(\bar{m}, m) - \quad /1/$$

варіаційний ряд зрізаної вибірки з абсолютно неперервної популяції ξ , а

$$y(\bar{1}, n) \leq \dots \leq y(\bar{j}, n) \leq \dots \leq y(\bar{n}, n) - \quad /2/$$

варіаційний ряд зрізаної вибірки з абсолютно неперервної популяції η . Для означеності, без обмеження загальності, можна вважати, що $m \leq n$. Якщо числа /1/ і /2/ позначають напрацювання до відмови F або зупинки S технічної одиниці, то в разі відмови числа $\bar{l}$ та $\bar{j}$ виражають середній ранг цієї відмови. На основі варіаційних рядів /1/ і /2/ знаходимо частотні емпіричні функції розподілу $F_m(x)$ та $G_n(x)$, причому в моменти відмов

$$F_m(x(\bar{l}, m)) = \frac{\bar{l}}{m}, \quad G_n(y(\bar{j}, n)) = \frac{\bar{j}}{n}. \quad /3/$$

Потрібно перевірити гіпотезу про те, що популяції ξ та η стохастично еквівалентні.

$$H: E F_m(x) \equiv E G_n(x), \quad -\infty < x < \infty. \quad /4/$$

За міру відхилення між вибірками /1/ та /2/ приймаємо статистику

$$D_{mn} = \sup_{-\infty < x < \infty} |F_m(x) - G_n(x)|. \quad /5/$$

Нехай α - заданий рівень значущості при $0 < \alpha \leq 0,05$, $D_{mn}(\alpha)$ - критичне значення статистики D_{mn} , визначене як розв'язок нерівностей

$$P\{D_{mn} \geq D_{mn}(\alpha)\} \leq \alpha, \quad P\{D_{mn} < D_{mn}(\alpha)\} > 1 - \alpha. \quad /6/$$

Якщо емпіричне значення статистики D_{mn} не менше від $D_{mn}(\alpha)$, то на рівні значущості α гіпотезу /4/ відкидаємо. Критичні значення для перевірки однорідності двох вибірок /критерій Смирнова/ табульовані* при $n = 1/1/20$, $m = 1/1/n$ та $\alpha = 0,10; 0,05; 0,02; 0,01$. При $n > 20$ для обчислення $D_{mn}(\alpha)$ слід користуватися наближенням

$$D_{mn}(\alpha) \approx D_{\frac{m+n}{mn}}(\alpha) + \frac{1}{k(m,n)} + \frac{1}{3n} - \frac{(n+1)[m-d(m,n)]}{2mn[m+n+d(m,n)]}, \quad /7/$$

де значення $D_{\frac{m+n}{mn}}(\alpha)$ можна визначити інтерполяцією в таблиці /див.

Большев Л., Смирнов Н., с.3 / для критичних значень при критерію Колмогорова, $k(m,n)$ - найменше спільне кратне, а $d(m,n)$ - найбільший спільний дільник чисел m і n .

П р и к л а д. Дано дві незалежні вибірки напрацьовань до відмови F або зупинки S технічної одиниці

(x): 2175 F 3400 S 4390 F 5242 F 6010 S
6799 F 7557 F 8326 F 9000 S 9000 S /8/
10894 F 11942 F 12000 S 14845 F 17581 F

та

(y): 2000 S 4019 F 5199 F 6264 F 7000 S
8325 F 9413 F 10608 F 11000 S 13795 F /9/
16704 F

Чи можна прийняти гіпотезу про однорідність вибірок /8/ і /9/ ?

Подано спільний варіаційний ряд (x) та (y) для моментів відмов у вибірках /8/ і /9/, відповідні середні ранги відмов $\bar{I}$ та $\bar{J}$ відповідно значення F_{15} та G_{11} частотних емпіричних функцій розподілу в пунктах спільного варіаційного ряду, а також модулі різниць відповідних значень $|F_{15} - G_{11}|$:

(x)	$\bar{I}$	(y)	$\bar{J}$	F_{15}	G_{11}	$F_{15} - G_{11}$
2157	I			0,06666	0	0,06666
		4019	1,09091	0,06666	0,09917	0,03251
4390	2,07143			0,13809	0,09917	0,03892
		5199	2,18182	0,13809	0,19835	0,06025
5242	3,04286			0,20952	0,19835	0,01118
		6264	3,27273	0,20952	0,29752	0,08800
6799	4,31169			0,28745	0,29752	0,01007

* Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М., 1963. С.350-352.

7557	5,48052			0,36537	0,29752	0,06785
		8325	4,51948	0,36537	0,41086	0,04549
8326	6,64935			0,44229	0,41086	0,03243
		9413	5,76623	0,44329	0,52420	0,00091
		10608	7,01299	0,44329	0,63754	0,19425
I0894	8,20779			0,54719	0,63754	0,09036
II942	9,76623			0,65108	0,63754	0,13538
		I3795	8,67532	0,65108	0,78867	0,13753
I484I	II,844I5			0,7896I	0,78867	0,00094
		I6704	10,33766	0,7896I	0,93979	0,15018
I758I	13,92207			0,92814	0,93979	0,01165

Найбільше значення в останній графі дорівнює 0,19425. З таблиці /див. Большев Л., Смирнов Н., с. 351/ знаходимо критичне значення $D_{II,15} / 0,05 / = 0,5071895$. Зазначимо, що за формулою /7/ дістанемо приблизно таке ж число. Гіпотезу про однорідність вибірок /8/ і /9/ приймаємо.

З а у в а ж е н н я. За цією ж таблицею /с.350-352/ для знаходження критичних значень статистики Смирнова припускається повнота вибірок. Не можна користуватися у випадку слабо зрізаних, а також у разі сильно зрізаних вибірок, якщо гіпотеза відхиляється.

Стаття надійшла до редколегії 14.03.91