УДК 519.2:303.5(168.2+004.421.2)

Ф. В. МОЦНИЙ
доктор фізико-математичних наук, професор,
завідувач кафедри прикладної математики,
Національна академія статистики, обліку та аудиту

Сучасний базовий інструментарій математичної статистики
Частина ІІ

Ця стаття – оглядова. У ній узагальнено результати численних досліджень сучасного базового інструментарію математичної статистики. Розглянуто основні поняття і вибіркові характеристики масових випадкових явищ з урахуванням їх особливостей. Розв'язано типові задачі. Проілюстровано теорему Гливенка – Кантеллі. Оцінено параметри статистичних розподілів.

Ключові слова: *генеральна сукупність, вибірка, варіаційний ряд, розподіл вибірки, полігон частот, гістограма частот, функція розподілу, крива розподілу, теорема Гливенка – Кантеллі, оцінка, дисперсія, середнє квадратичне відхилення, статистичний розподіл.*

Частина ІІ. Вибіркові характеристики статистичних розподілів

Числові характеристики випадкових величин: *медіана, мода, математичне сподівання, дисперсія, середнє квадратичне відхилення, початкові і центральні моменти різних порядків* розглянуто в [19]. Аналогічні числові характеристики мають місце для статистичних розподілів. Ці дані можна використати при досліджені та встановлені законів розподілу.

2.1. Одномірний дискретний статистичний розподіл вибірки

Медіана m_e^* – варіанта вибірки, що поділяє варіаційний ряд на дві рівні частини за кількістю варіант.

Мода m_o^* – варіанта вибірки, яка має найбільшу частоту появи. Так, для дискретного статистичного розподілу вибірки, заданого в табл. 2 (ч. І), мода дорівнює 10, бо її частота появи є найбільшою і дорівнює 50.

Вибіркове середнє – статистична оцінка математичного сподівання, яка визначається середнім арифметичним значенням ознаки вибіркової сукупності. Якщо всі варіанти (вибіркові значення) x_i варіаційного ряду з'являються у вибірці з частотою n_i, то *вибіркове середнє* визначається за формулою:

$$\bar{x}_s = \frac{1}{n} \cdot \sum_{i=1}^{m} x_i \cdot n_i \tag{11}$$

де $n = \sum_{i=1}^{m} n_i$ – обсяг вибірки, m – кількість варіант. Характерно, що вибірковим середнім притаманна властивість стабільності. Так, якщо з однієї і тієї ж генеральної сукупності взяти кілька вибірок великого обсягу, то їхні вибіркові середні будуть приблизно рівними.

Вибіркова дисперсія – статистична оцінка генеральної дисперсії, яка визначається середнім арифметичним квадратів відхилень варіант x_i від вибіркового середнього $\bar{x}_s$. Якщо варіанти x_i варіаційного ряду з'являються у вибірці з частотою n_i, то *вибіркова дисперсія* визначається таким чином:

$$D_s = \frac{\sum_{i=1}^{m} (x_i - \bar{x}_s)^2 \cdot n_i}{n}, \tag{12}$$

або

$$D_s = \frac{\sum_{i=1}^{m} (x_i)^2 \cdot n_i}{n} - (\bar{x}_s)^2. \tag{13}$$

Отже, вибіркова дисперсія дорівнює різниці між середнім квадратів значень ознаки і квадратом вибіркового середнього.

Для характеристики розсіювання значень ознаки вибірки навколо їхнього середнього значення окрім дисперсії використовують також зведену числову характеристику – *вибіркове середнє квадратичне відхилення*, розмірність якого співпадає з розмірністю випадкової величини.

Вибіркове середнє квадратичне відхилення – квадратний корінь із вибіркової дисперсії.

$$\sigma_S = \sqrt{D_S} \, . \tag{14}$$

Зауваження.

1. Якщо за статистичну оцінку генеральної сукупності взяти вибіркову дисперсію D_S, то ця оцінка приведе до систематичних похибок, наслідком яких буде більш низьке значення генеральної дисперсії D_g. Це пояснюється тим, що D_S є зміщеною крапковою (точковою) статистичною оцінкою для D_g. Дійсно, зв'язок математичного сподівання вибіркової дисперсії і дисперсії генеральної сукупності визначається формулою:

$$M(D_S) = \frac{n-1}{n} \cdot D_g \, , \tag{15}$$

де $\dfrac{n-1}{n}$ – коефіцієнт зміщення.

2. Якщо помножити формулу (12) на коефіцієнт зміщення $\dfrac{n}{n-1}$, отримаємо *виправлену дисперсію:*

$$S^2 = \frac{n}{n-1} \cdot D_S = \frac{n}{n-1} \cdot \frac{\sum\limits_{i=1}^{m} (x_i - \bar{x}_S)^2 \cdot n_i}{n} = \frac{\sum\limits_{i=1}^{m} (x_i - \bar{x}_S)^2 \cdot n_i}{n-1} \, . \tag{16}$$

3. *Виправлена дисперсія* S^2 є незміщеною статистичною оцінкою генеральної дисперсії D_g, оскільки

$$M(S^2) = M\left(\frac{n}{n-1} \cdot D_S\right) = \frac{n}{n-1} \cdot M(D_S) = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot D_g = D_g .$$

4. *Виправлена дисперсія* S^2 є також спроможною статистичною оцінкою генеральної дисперсії D_g за означенням (S^2 збігається по ймовірності до оцінювального параметра D_g при $n \to \infty$). Якщо порівняти формули (12) і (16), помічаємо, що при великих обсягах вибірки вибіркова і виправлена дисперсії відрізняються мало. Тому вже при $n < 30$ за дисперсію D_g генеральної сукупності приймається виправлена дисперсія S^2.

Виправлене середнє квадратичне відхилення обчислюється за формулою:

$$S = \sqrt{\frac{n}{n-1} \cdot D_S} \, . \tag{17}$$

При цьому слід наголосити, що *виправлене середнє квадратичне відхилення* є зміщеною точковою статистичною оцінкою для *середнього квадратичного відхилення*

σ_g генеральної сукупності і визначається так:

$$M(S)=\sqrt{\frac{2}{k}}\cdot\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\cdot\sigma_g \quad . \tag{18}$$

Тут $k=(n-1)$ – кількість степенів свободи; n – обсяг вибірки;

$\sqrt{\frac{2}{k}}\cdot\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}$ – коефіцієнт зміщення;

$\Gamma(\frac{k+1}{2})$ і $\Gamma(\frac{k}{2})$ – гамма функції $(\Gamma(x)=\int_0^{\infty}t^{x-1}e^{-t}dt,$ причому $\Gamma(x)=(x-1)!,$ якщо $x=1, 2, 3, 4,...$).

2.2. Одномірний інтервальний статистичний розподіл вибірки

Медіана одномірного інтервального статистичного розподілу вибірки обчислюється таким чином [8]:

$$m_e^{in}=x_{i-1}+\frac{0{,}5-F_n^*(x_{i-1})}{F_n^*(x_i)-F_n^*(x_{i-1})}\cdot l \ , \tag{19}$$

де $F_n^*(x_{i-1})$ і $F_n^*(x_i)$ – значення емпіричної функції розподілу вибірки варіант x_{i-1} та x_i відповідно; x_{i-1} – початок *медіанного інтервалу* $[x_{i-1}, x_i]$,тобто інтервалу, всередині якого обов'язково знайдеться таке значення варіанти $X=m_e$, при якому $F_n^*(m_e)=\frac{1}{2}$; $l=(x_i-x_{i-1})$ – величина частинного інтервалу.

Мода одномірного інтервального статистичного розподілу вибірки знаходиться у такий спосіб [8]:

$$m_o^{in}=x_{i-1}^*+\frac{n_{m_o}-n_{m_o-1}}{(n_{m_o}-n_{m_o-1})+(n_{m_o}-n_{m_o+1})}\cdot l \ , \tag{20}$$

Тут x_{i-1}^* – початок *модального інтервалу* $[x_{i-1}^*, x_i^*]$, тобто інтервалу, всередині якого варіанта $X=m_o$ вибірки має найбільшу частоту появи; n_{m_o} – частота модального інтервалу; n_{m_o-1} – частоти домодального інтервалу; n_{m_o+1} – частоти післямодального інтервалу; $l=(x_i^*-x_{i-1}^*)$ – величина частинного інтервалу.

Зауважимо, що *медіанний і модальний* інтервали статистичного розподілу вибірки можуть потрапити як в один, так і в різні інтервали. Тому початок модального інтервалу позначено зірочкою, щоб відрізнити його від початку медіанного інтервалу.

Обчислимо медіану і моду *для парного і непарного інтервалів*, коли обидва інтервали знаходяться в одному місці, або в різних місцях.

Приклад 4. За інтервальним статистичним розподілом вибірки, заданим в табл. 1 (ч. І), знайти емпіричну функцію $F_n^*(x)$ розподілу вибірки, побудувати її графік та обчислити m_e^{in} і m_o^{in} .

Розв'язання.

Функція $F_n^*(x)$ має вигляд:

$$F_n^*(x)=\begin{cases}0; & x\le 1.\\ W_1=0{,}0182; & 1<x\le 3.\\ W_1+W_2=0{,}0182+0{,}0909=0{,}1091; & 3<x\le 5.\\ W_1+W_2+W_3=0{,}1091+0{,}1636=0{,}2727; & 5<x\le 7.\\ W_1+W_2+W_3+W_4=0{,}2727+0{,}2364=0{,}5091; & 7<x\le 9.\\ W_1+W_2+W_3+W_4+W_5=0{,}5091+0{,}2909=0{,}8; & 9<x\le 11.\\ W_1+W_2+W_3+W_4+W_5+W_6=0{,}8+0{,}1818=0{,}9818; & 11<x\le 13.\\ W_1+W_2+W_3+W_4+W_5+W_6+W_7=0{,}9818+0{,}0182=1; & 13<x\le 15.\end{cases}$$

Графік цієї функції показано на рис. 6. Із цього графіка і табл. 1 (ч. І) помічаємо,

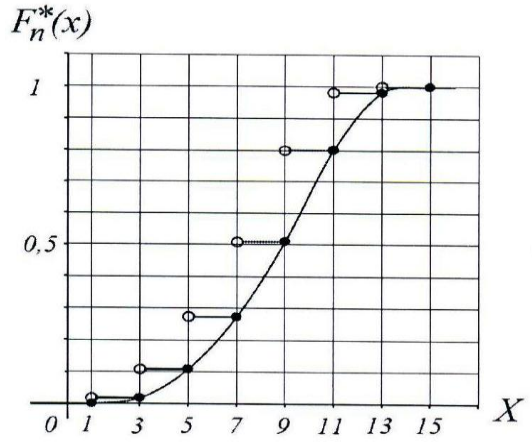

Рис. 6. Графік емпіричної функції $F_n^*(x)$ інтервальної вибірки, заданої в табл. 1 (ч. І)

що медіанний і модальний інтервали співпадають ([7, 9]) і є парними. Для медіанного інтервалу маємо:

$x_{i-1}=7$; $x_i=9$; $l=2$; $F_n^*(7)=0{,}2727$; $F_n^*(9)=0{,}5091$. Підставимо знайдені числа у формулу (19), отримаємо:

$$m_e^{in}=7+\frac{0{,}5-0{,}2727}{0{,}5091-0{,}2727}\cdot 2\approx 8{,}92.$$

Використаємо ці ж дані для модального інтервалу, знайдемо, що $x_{i-1}^*=7$; $x_i^*=9$; l=2; $n_{m_o}=16$; $n_{m_o-1}=13$; $n_{m_o+1}=10$. Підставимо ці величини у формулу (20), отримаємо:

$$m_o^{in}=7+\frac{16-13}{(16-13)+(16-10)}\cdot 2=7+\frac{3}{9}\approx 7{,}33.$$

Приклад 5. За інтервальним статистичним розподілом вибірки, заданим у таблиці, знайти m_e^{in} і m_o^{in}.

Частинний інтервал	9-12	12-15	15-18	18-21
n_i	10	18	14	8
W_i	0,20	0,36	0,28	0,16

Розв'язання.

Обсяг вибірки n=50. Контроль: 0,20+0,36+0,28+0,16=1 (вибірку як статистичний ряд складено вірно). Тут медіанний і модальний інтервали співпадають [12, 15], але є непарними.

Медіанний інтервал [12, 15]: $x_{i-1}=12$; $x_i=15$; $l=3$; $F_n^*(12)=0,20$; $F_n^*(15)=0,56$.

Підставимо ці величини у формулу (19), отримаємо:

$$m_e^{in}=12+\frac{0,5-0,20}{0,56-0,20}\cdot 3=14,5.$$

Модальний інтервал [12, 15]: $x_{i-1}^*=12$; $x_i^*=15$; $l=3$; $n_{m_o}=18$; $n_{m_o-1}=10$; $n_{m_o+1}=14$. Підставимо ці величини у формулу (20), отримаємо:

$$m_o^{in}=12+\frac{18-10}{(18-10)+(18-14)}\cdot 3=14.$$

Приклад 6. За інтервальним статистичним розподілом вибірки, заданим у таблиці, знайти m_e^{in} і m_o^{in}.

Частинний інтервал	0-4	4-8	8-12	12-16
n_i	12	15	17	8
w_i	0,23	0,29	0,33	0,15

Розв'язання.

Обсяг вибірки n=52. Контроль: 0,23+0,29+0,33+0,15=1 (вибірку як статистичний ряд складено вірно). Тут медіанний і модальний інтервали парні, але різні: [4, 8] і [8, 12] відповідно.

Медіанний інтервал [4, 8]: $x_{i-1}=4$; $x_i=8$; $l=4$; $F_n^*(4)=0,23$; $F_n^*(15)=0,52$.

Підставимо ці числа у формулу (19), отримаємо:

$$m_e^{in}=4+\frac{0,5-0,23}{0,52-0,23}\cdot 4=7,72.$$

Модальний інтервал [8, 12]: $x_{i-1}^*=8$; $x_i^*=12$; $l=4$; $n_{m_o}=17$; $n_{m_o-1}=15$; $n_{m_o+1}=8$. Підставимо ці числа у формулу (20), знайдемо:

$$m_o^{in}=8+\frac{17-15}{(17-15)+(17-8)}\cdot 4\approx 8,73.$$

Для обчислення *вибіркових середнього* $\bar{x}_s^{in}$, *дисперсії* D_s^{in} і *середнього квадратичного відхилення* σ_s^{in} інтервального статистичного розподілу вибірки треба перейти від *інтервального розподілу* вибірки до *дискретного*, прийнявши за варіанти середини часткових інтервалів: $x_i^{mid}=x_{i-1}+\frac{l}{2}=x_i-\frac{l}{2}$ (див. табл. 1 (ч. I)). Тоді $\bar{x}_s^{in}$, D_s^{in} і σ_s^{in} визначатимуться за формулами:

$$\bar{x}_s^{in} = \frac{1}{n} \cdot \sum_{i=1}^{n^*} x_i^{mid} \cdot n_i \qquad (21)$$

$$D_s^{in} = \frac{\sum_{i=1}^{n^*} (x_i^{mid})^2 \cdot n_i}{n} - (\bar{x}_s^{in})^2 ; \qquad (22)$$

$$\sigma_s^{in} = \sqrt{D_s^{in}} \qquad (23)$$

відповідно, де n^* – кількість часткових інтервалів; $n = \sum_{i=1}^{n^*} n_i$ – обсяг вибірки.

Приклад 7. За даними табл. 1 (ч. І) інтервального статистичного розподілу вибірки знайти $\bar{x}_s^{in}$, D_s^{in} і σ_s^{in}.

Розв'язання.

Тут n^* =7. За формулами (21) – (23) отримаємо:

$$\bar{x}_s^{in} = \frac{2 \cdot 1 + 4 \cdot 5 + 6 \cdot 9 + 8 \cdot 13 + 10 \cdot 16 + 12 \cdot 10 + 14 \cdot 1}{55} \approx 8,62.$$

$$D_s^{in} = \frac{2^2 \cdot 1 + 4^2 \cdot 5 + 6^2 \cdot 9 + 8^2 \cdot 13 + 10^2 \cdot 16 + 12^2 \cdot 10 + 14^2 \cdot 1}{55} - (8,62)^2 \approx 7,08.$$

$$\sigma_s^{in} = \sqrt{7,08} \approx 2,66.$$

2.3. Двомірний дискретний статистичний розподіл вибірки

Перелік варіант $Y = y_i$ і $X = x_j$ та відповідних їм частот n_{ij} спільної появи називається *двомірним дискретним статистичним розподілом вибірки*, елементам якої притаманні кількісні ознаки X і Y (див. табл. 4).

Таблиця 4

Двомірний дискретний статистичний розподіл вибірки

$Y = y_i$	$X = x_J$					
	x_1	x_2	x_3	…	x_m	n_{y_i}
y_1	n_{11}	n_{12}	n_{13}	…	n_{1m}	n_{y_1}
y_2	n_{21}	n_{22}	n_{23}	…	n_{2m}	n_{y_2}
y_3	n_{31}	n_{32}	n_{33}	…	n_{3m}	n_{y_3}
…	…	…	…	…	…	…
y_k	n_{k1}	n_{k2}	n_{k3}	…	n_{km}	n_{y_k}
n_{x_j}	n_{x_1}	n_{x_2}	n_{x_3}	…	n_{x_m}	

Зазначимо, що $n_{y_i} = \sum_{j=1}^{m} n_{ij}$; $n_{x_j} = \sum_{i=1}^{k} n_{ij}$; $n = \sum_{i=1}^{k} \sum_{J=1}^{m} n_{ij} = \sum_{i=1}^{k} n_{y_i} = \sum_{J=1}^{m} n_{x_j}$.

Загальні середня величина, дисперсія і середнє квадратичне відхилення ознаки X визначаються, відповідно, за формулами:

$$\bar{x} = \frac{\sum_{i=1}^{k} \sum_{J=1}^{m} x_j \cdot n_{ij}}{n} = \frac{\sum_{J=1}^{m} x_j \cdot n_{x_j}}{n} ; \tag{24}$$

$$D_x = \frac{\sum_{i=1}^{k} \sum_{j=1}^{m} (x_j)^2 \cdot n_{ij}}{n} - (\bar{x})^2 = \frac{\sum_{j=1}^{m} (x_j)^2 \cdot n_{x_j}}{n} - (\bar{x})^2 ; \tag{25}$$

$$\sigma_x = \sqrt{D_x}. \tag{26}$$

Загальні середня величина, дисперсія і середнє квадратичне відхилення ознаки Y визначаються, відповідно, таким чином:

$$\bar{y} = \frac{\sum_{i=1}^{k} \sum_{J=1}^{m} y_i \cdot n_{ij}}{n} = \frac{\sum_{i=1}^{k} y_i \cdot n_{y_i}}{n} ; \tag{27}$$

$$D_y = \frac{\sum_{i=1}^{k} \sum_{j=1}^{m} (y_i)^2 \cdot n_{ij}}{n} - (\bar{y})^2 = \frac{\sum_{i=1}^{k} (y_i)^2 \cdot n_{y_i}}{n} - (\bar{y})^2 ; \tag{28}$$

$$\sigma_y = \sqrt{D_y}. \tag{29}$$

Приклад 8. За заданим у таблиці дискретним статистичним розподілом вибірки

$Y = y_i$	$X = x_j$					
	5	10	15	20	25	n_{y_i}
1	1	3	2	4	5	15
3	3	4	5	3	5	20
5	1	3	7	8	11	30
7	5	5	6	10	9	35
n_{x_j}	10	15	20	25	30	

ознак X та Y обчислити $\bar{x}, \bar{y}, D_x, \sigma_x, D_y, \sigma_y$.

Розв'язання.

Тут n=100. Використаємо формули (24) – (29), отримаємо:

$$\bar{x} = \frac{5 \cdot 10 + 10 \cdot 15 + 15 \cdot 20 + 20 \cdot 25 + 25 \cdot 30}{100} = 17{,}5;$$

$$\bar{y}=\frac{1\cdot 15+3\cdot 20+5\cdot 30+7\cdot 35}{100}=4,7;$$

$$D_x=\frac{5^2\cdot 10+10^2\cdot 15+15^2\cdot 20+20^2\cdot 25+25^2\cdot 30}{100}-17,5^2=43,75;$$

$$\sigma_x=\sqrt{43,75}\approx 6,61.$$

$$D_y=\frac{1^2\cdot 15+3^2\cdot 20+5^2\cdot 30+7^2\cdot 35}{100}-4,7^2=4,51;$$

$$\sigma_y=\sqrt{4,51}\approx 2,12.$$

Зв'язок ознак X та Y у математичній статистиці називається кореляційним і обчислюється за формулою:

$$K_{xy}=\frac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m} y_i\cdot x_j\cdot n_{ij}}{n}-\bar{x}\cdot\bar{y}. \qquad (30)$$

Тут K_{xy} – *емпіричний кореляційний момент*. Якщо $K_{xy}\neq 0$, то кореляційний зв'язок ознак X *та* Y є. Якщо ж $K_{xy}=0$, то такого зв'язку немає.

Приклад 9. За дискретним статистичним розподілом вибірки ознак X і Y, заданим у таблиці прикладу 8, обчислити K_{xy} .

Розв'язання.

Використаємо формулу (30), тоді

$$\sum_{i=1}^{k}\sum_{j=1}^{m} y_i\cdot x_j\cdot n_{ij}=y_1\cdot x_1\cdot n_{11}+y_1\cdot x_2\cdot n_{12}+y_1\cdot x_3\cdot n_{13}+y_1\cdot x_4\cdot n_{14}+y_1\cdot x_5\cdot n_{15}+$$

$$+y_2\cdot x_1\cdot n_{21}+y_2\cdot x_2\cdot n_{22}+y_2\cdot x_3\cdot n_{23}+y_2\cdot x_4\cdot n_{24}+y_2\cdot x_5\cdot n_{25}+$$

$$+y_3\cdot x_1\cdot n_{31}+y_3\cdot x_2\cdot n_{23}+y_3\cdot x_3\cdot n_{33}+y_3\cdot x_4\cdot n_{34}+y_3\cdot x_5\cdot n_{35}+$$

$$+y_4\cdot x_1\cdot n_{41}+y_4\cdot x_2\cdot n_{42}+y_4\cdot x_3\cdot n_{43}+y_4\cdot x_4\cdot n_{44}+y_4\cdot x_5\cdot n_{45}=$$

$$=1\cdot 5\cdot 1+1\cdot 10\cdot 3+1\cdot 15\cdot 2+1\cdot 20\cdot 4+1\cdot 25\cdot 5+3\cdot 5\cdot 3+3\cdot 10\cdot 4+$$

$$+3\cdot 15\cdot 5+3\cdot 20\cdot 3+3\cdot 25\cdot 5+5\cdot 5\cdot 1+5\cdot 10\cdot 3+5\cdot 15\cdot 7+5\cdot 20\cdot 8+$$

$$+5\cdot 25\cdot 11+7\cdot 5\cdot 5+7\cdot 10\cdot 5+7\cdot 15\cdot 6+7\cdot 20\cdot 10+7\cdot 25\cdot 9=8220.$$

Остаточно $K_{xy}=\frac{8220}{100}-17,5\cdot 4,7\approx 82,20-82,25\approx -0,05$. Оскільки $K_{xy}\approx -0,05\neq 0$, то ознаки X та Y зв'язані між собою від'ємним кореляційним зв'язком.

2.4. Двомірний дискретний умовний статистичний розподіл вибірки

Перелік варіант ознаки $Y=y_i$ та відповідних їм частот n_{ij}, узятих при фіксованому значенні ознаки n_{ij} , називається *двомірним дискретним умовним ста-*

тистичним розподілом вибірки ознаки Y при фіксованому значенні ознаки $X = x_j$, тобто $Y/X = x_j$.

Перелік варіант ознаки $X = x_j$ та відповідних їм частот n_{ij}, узятих при фіксованому значенні ознаки $Y = y_i$, називається *двомірним дискретним умовним статистичним розподілом вибірки ознаки X* при фіксованому значенні ознаки $Y = y_i$, тобто $X\ /Y = y_i$.

Умовні середня величина, дисперсія та середнє квадратичне відхилення ознак Y та X, відповідно, такі:

$$\bar{y}_{X=x_J} = \frac{\sum\limits_{i=1}^{k} y_i \cdot n_{ij}}{\sum\limits_{i=1}^{k} n_{ij}} = \frac{\sum\limits_{i=1}^{k} y_i \cdot n_{y_i}}{n_{x_j}}\ ; \tag{31}$$

$$D(Y/X = x_j) = \frac{\sum\limits_{i=1}^{k} (y_i)^2 \cdot n_{ij}}{n_{x_j}} - (\bar{y}_{X=x_j})^2\ ; \tag{32}$$

$$\sigma(Y/X = x_j) = \sqrt{D(Y/X = x_j)}\ ; \tag{33}$$

$$\bar{x}_{Y=y_i} = \frac{\sum\limits_{j=1}^{m} x_j \cdot n_{ij}}{\sum\limits_{j=1}^{m} n_{ij}} = \frac{\sum\limits_{j=1}^{m} x_j \cdot n_{ij}}{n_{y_i}}\ ; \tag{34}$$

$$D(X/Y = y_i) = \frac{\sum\limits_{j=1}^{m} (x_j)^2 \cdot n_{ij}}{n_{y_i}} - (\bar{x}_{Y=y_i})^2\ ; \tag{35}$$

$$\sigma(X/Y = y_i) = \sqrt{D(X/Y = y_i)}\ . \tag{36}$$

Приклад 10. За дискретним статистичним розподілом вибірки ознак X і Y, заданим у таблиці прикладу 8, побудувати умовні статистичні розподіли $Y/X = 10$; $Y/X = 5$ та обчислити умовні числові характеристики.

Розв'язання.

Оскільки $Y/X = 10$, то двомірний дискретний умовний статистичний розподіл вибірки згідно з означенням є таким:

$Y = y_i$	1	3	5	7
n_{i2}	3	4	3	5

Використаємо формули (31) – (33), знайдемо, відповідно, умовні середню величину, дисперсію та середнє квадратичне відхилення ознаки Y.

$$\bar{y}_{X=10} = \frac{1\cdot3+3\cdot4+5\cdot3+7\cdot5}{100} = 0{,}65;$$

$$D(Y/X=10) = \frac{1^2\cdot3+3^2\cdot4+5^2\cdot3+7^2\cdot5}{100} - 0{,}65^2 \approx 3{,}23;$$

$$\sigma(Y/X=10) = \sqrt{3{,}23} \approx 1{,}8.$$

Оскільки $Y/X=5$, то двомірний дискретний умовний статистичний розподіл вибірки згідно з його означенням запишемо таким чином:

$X=x_j$	5	10	15	20	25
n_{3j}	1	3	7	8	11

Використаємо формули (34) – (36), знайдемо, відповідно, умовні середню величину, дисперсію та середнє квадратичне відхилення ознаки *X*.

$$\bar{x}_{Y=5} = \frac{5\cdot1+10\cdot3+15\cdot7+20\cdot8+25\cdot11}{100} = 5{,}75.$$

$$D(X/Y=5) = \frac{5^2\cdot1+10^2\cdot3+15^2\cdot7+20^2\cdot8+25^2\cdot11}{100} - 5{,}75^2 \approx 86{,}69.$$

$$\sigma(X/Y=5) = \sqrt{86{,}69} \approx 9{,}31.$$

Отже, у цій статті узагальнено численні результати сучасних досліджень базового інструментарію математичної статистики. Робота може бути корисна студентам, аспірантам, фахівцям в галузі статистики та іншим спеціалістам при вивчені курсу "Теорія ймовірності і математична статистика", а також при досліджені та встановлені законів статистичних розподілів.

Висновки

1. Вибіркова дисперсія D_s є зміщеною точковою статистичною оцінкою дисперсії D_g генеральної сукупності, що призводить до систематичних похибок і зменшення значення D_g.

2. Виправлена дисперсія S^2 є незміщеною спроможною статистичною оцінкою генеральної дисперсії D_g. При великих обсягах вибірки виправлена і вибіркова дисперсії відрізняються мало, і тому вже при $n<30$ за дисперсію генеральної сукупності приймається виправлена дисперсія.

3. Виправлене середнє квадратичне відхилення S є зміщеною точковою статистичною оцінкою для середнього квадратичного відхилення σ_g генеральної сукупності.

4. Для знаходження вибіркових середнього $\bar{x}_s^{in}$, дисперсії D_s^{in} і середнього квадратичного відхилення D_s^{in} інтервального статистичного розподілу вибірки необхідно перейти від інтервального розподілу вибірки до дискретного, прийнявши за варіанти середини часткових інтервалів.

5. Двомірний дискретний статистичний розподіл вибірки визначається переліком варіант $Y=y_i$ і $X=x_j$ та відповідних їм частот n_{ij} спільної появи. Загальні середня величина, дисперсія і середнє квадратичне відхилення ознак *X* і *Y* обчислюються за формулами (24) – (26) і (27) – (29) відповідно.

6. Кореляційний зв'язок ознак *X* і *Y* визначається емпіричним кореляційним моментом K_{xy}. Якщо $K_{xy}\neq0$, то кореляційний зв'язок між ознаками *X* та *Y* є. Якщо ж $K_{xy}=0$, то такого зв'язку немає.

7. Двомірний дискретний умовний статистичний розподіл вибірки визначається переліком варіант ознаки $Y = y_i$ та відповідних їм частот n_{ij} при фіксованому значенні ознаки $X = x_j$, переліком варіант $X = x_j$ та відповідних їм частот n_{ij} при фіксованому значенні ознаки $Y = y_i$. Умовні середня величина, дисперсія та середнє квадратичне відхилення ознак X і Y обчислюються за формулами (31) – (36) відповідно.

Дані про числові характеристики в одномірному і двомірному випадках статистичного розподілу вибірки можна використати для вивчення і встановлення законів статистичних розподілів.

Список використаних джерел

1. Glivenko V. I. Sulla determinazione empirica della legge di probabilita / V. I. Glivenko. – Giorn. Ist. Ital. Attuari. – 1933. – Vol. 4. – P. 92–99.
2. Cantelli F. P. Sulla determinazione empirica delle leggi di probabilita / F. P. Cantelli. – Giorn. Ist. Ital. Attuari. – 1933. – Vol. 4. – P. 421–424.
3. Ілюстрація теореми Гливенка-Кантеллі [Електронний ресурс]. – Режим доступу : //http://www.wolfram. com/mathematica/new-in-8/probability-and-statistics-solvers-and-roperties/demonstrate-the-glivenko-cantelli-theorem.ru.html.
4. Kolmogorov A. N. Sulla determinazione empirica di una legge di distributione / A. N. Kolmogorov. – Giorn. Inst. Ital. Attuari. – 1933. –Vol. 4. – P. 83–91.
5. Snedecor G. W. Statistical Methods : [6th edition] / G. W. Snedecor, W. G. Cochran. – Iowa : The Iowa State University Press, AMES, 1971. – 593 p.
6. Spurr. W. A. Statistical Analysis for Business Decisions : [9th edition] / W. A. Spurr, Ch. P. Bonini. – Illinois : Richard D. Irwin, Inc., 1977. – 724 p.
7. Гмурман В. Е. Теория вероятностей и математическая статистика : [учебн. пособ.] / В. Е. Гмурман. – М. : Высшая школа,1999. – 400 с.
8. Жлуктенко В. І. Теорія ймовірностей і математична статистика : [у 2-х ч.]: Ч. 2. : Математична статистика : [навч. посіб.] / В. І. Жлуктенко, С. І. Наконечний, С. С. Савіна. – К. : КНЕУ, 2001. – 336 с.
9. Сеньо П. С. Теорія ймовірностей та математична статистика : [підручник] / П. С. Сеньо. – К. : Знання, 2007. – 556 с.
10. Пономаренко В. С. Багатовимірний аналіз соціально-економічних систем : [навч. посіб.] / В. С. Пономаренко, Л. М. Малярець. – Харків : ХНЕУ, 2009. – 384 с.
11. Dragulescu A. Statistical Mechanics of Money / A. Dragulescu, V. M. Yakovenko. – Eur. Phys. J. B., 2000. – Vol. 17. – P.723–729.
12. Acerbi C. Risk Aversion and Coherent Risk Measures: a Spectral Representation Theorem / C. Acerbi. – arXiv:cond-mat/0107190v1, 2001. – P. 1–11.
13. Yakovenko V. M. Colloquium: Statistical Methods of Money, Wealth and Income / V. M. Yakovenko, J. B. Rosser. – arXiv:0905.1518v2[q-finST], 2009. – P. 1–25.
14. Liang J. Valuation of Credit Default Swap with Counterparty Default Risk by Structural Model / J. Liang, P. Zhou, Y. Zhou, J. Ma. – Applied Mathematics. – 2011. – Vol. 2. – P. 106–117.
15. Кулинич Р. О. Методика оцінювання стійкості обмінного курсу валют / Р. О. Кулинич. – Науковий вісник НАСОА. – 2014. – № 2. – С. 14–18.
16. Бронштейн И. Н. Справочник по математике для инженеров и учащихся вузов / И. Н. Бронштейн, К. А. Семендяев. – М. : Наука, 1986. – 544 с.
17. Bond M. E. Students' Perceptions of Statistics: an Exploration of Attitudes, Conceptualizations, and Content Knowledge of Statistics / M. E. Bond, S. N. Perkins, C. Ramirez. – Statistics Education Research Journal. – 2012. – Vol. 11 (2). – P. 6–25.
18. Dierdorp A. Meaningful Statistics in Professional Practices as a Bridge between Mathematics and Science: and Evaluation of a Design Research Project / A. Dierdorp, A. Bakker, Jan A. van Maanen, H. MC Eijkelhof. – International Journal of STEM Education. – 2014. – Vol. 1. – P. 1–15.
19. Моцний Ф. В. Курс лекцій з теорії ймовірностей : [навч. посіб., видання друге, Гриф МОНУ] / Ф. В. Моцний. – К. : ДП ІАА, 2013. – 205 с.

Ф. В. МОЦНЫЙ
доктор физико-математических наук, профессор,
заведующий кафедрой прикладной математики,
Национальная академия статистики, учета и аудита

Современный базовый инструментарий математической статистики

Эта статья – обзорная. В ней обобщены результаты многочисленных исследований современного базового инструментария математической статистики. Рассмотрены основные понятия и выборочные характеристики массовых случайных явлений с учетом их особенностей. Решены типичные задачи. Проиллюстрирована теорема Гливенко – Кантелли. Оценены параметры статистических распределений.

Ключевые слова: *генеральная совокупность, выборка, вариационный ряд, распределение выборки, полигон частот, гистограмма частот, функция распределения, кривая распределения, теорема Гливенко – Кантелли, оценка, дисперсия, среднее квадратическое отклонение, статистическое распределение.*

F. V. MOTSNYI
Dr. Sc. (Phys. & Math.), Professor,
Head of Department for Applied Mathematics,
National Academy of Statistics, Accounting and Audit

Advanced Base Tools of Mathematical Statistics

Mathematical statistics reflects the quantitative data of the experiments of the mass accidental events. It combines different methods of studies that directed at setting up regularities and producing scientific and practical conclusions. The base for these methods is mathematics. The character of objects studied is not essential for it. Therefore mathematical methods are used extensively to study statistical data. This is a review paper. The results of numerous experiments of present-day base tools of mathematical statistics are summed up. Key concepts and sample characteristics of mass accidental events are considered with take into account their specific features. The typical problems are solved. Glivenko – Cantelli theorem is illustrated by graphs. Statistical distributions parameters are estimated.

Keywords: *general combination, sample, variational series, sample distribution, frequency polygon, frequency histogram, function of distribution, distribution curve, Glivenko – Cantelli theorem, estimation, dispursion, standard deviation, statistical distribution.*