

Щехорський Анатолій

*кандидат фізико-математичних наук, доцент, доцент кафедри
математичного аналізу, бізнес-аналізу та статистики
Житомирський державний університет імені Івана Франка
Україна, м.Житомир*

МОДЕЛЮВАННЯ ПОКАЗНИКІВ МІНЛИВОСТІ В ОПИСОВІЙ СТАТИСТИЦІ

Дескриптивна статистика займається обробкою статистичних даних у вигляді таблиць і графіків, а також кількісним представленням за основними статистичними показниками зокрема, показниками мінливості. Показники мінливості не роблять висновок про генеральну сукупність, а лише про об'єкти вибірки і тому підлягають модельному процесу їх виникнення.

Мета статті. В літературних джерелах показники мінливості описової статистики мають емпіричне походження. На прикладі

коефіцієнта варіації запропоновано пояснити їх походження на основі математичного моделювання.

Виклад основного матеріалу. Для прикладу розглядається задача, в якій за розміром заробітних плат x_1, \dots, x_n і y_1, \dots, y_m співробітників двох фірм потрібно знайти показник найбільш ефективний для клієнта фірм. Такий показник повинен бути відносним безрозмірним і не перевищувати одиниці. На перший погляд такий показник можна побудувати наступним способом. Середній степінь (частка, доля) відхилення від середньої всіх зарплат в загальному відхиленні визначається за формулою

$$\frac{1}{n} \sum_{i=1}^n \frac{|x_i - \bar{x}|}{\sum_{i=1}^n |x_i - \bar{x}|}, \quad (1)$$

Формула (1) приводить до константи $\frac{1}{n}$ і тому не може слугувати відносним порівняльним показником. Середнє відхилення заробітної плати в загальному відхиленні однакове для всіх фірм за даним об'ємом статистичних даних.

Частка окремого лінійного відхилення в загальному відхиленні не приводить до бажаного показника. Для отримання потрібного показника у формулі (1) пропонується заміна загального відхилення $\sum_{j=1}^n (x_j - \bar{x})^2$ на степеневу функцію, близьку до лінійної, $\varphi(|x_i - \bar{x}|) = |x_i - \bar{x}|^\alpha$ ($\alpha > 1$). Після відповідних перетворень формула (1) матиме наступний вигляд

$$\sum_{j=1}^n \frac{|x_j - \bar{x}|}{n \sum_{u=1}^m \varphi(|x_i - \bar{x}|)} = \sum_{j=1}^n \frac{|x_j - \bar{x}|}{\frac{n^2 |x_i - \bar{x}|^\alpha}{n}} \quad \text{Приведемо знаменник до лінійних одиниць}$$

$$\text{виміру, } \sum_{j=1}^n \frac{|x_j - \bar{x}|}{\sqrt[\alpha]{\frac{n^2 |x_i - \bar{x}|^\alpha}{n}}} = \sum_{i=1}^n \frac{\frac{|x_j - \bar{x}|}{n^{\frac{2}{\alpha}}}}{\sqrt[\alpha]{\frac{|x_i - \bar{x}|^\alpha}{n}}}. \quad \text{За змістом, чисельник повинен бути}$$

середнім лінійним відхиленням. Це можливо тільки при $\alpha=2$. Отримаємо показник

$$\frac{1}{n} \sum_{i=1}^n \frac{|x_j - \bar{x}|}{\sigma} \quad (2)$$

де $\sigma = \sqrt{\frac{\sum_{i=1}^n |x_i - \bar{x}|^2}{n}}$ є середнім квадратичним відхиленням значень

ознаки

За формулою (2) отримано показник, як степінь (частка, доля) середнього відхилення ознаки (зарплати) в загальному її відхиленні (відхиленні заробітних плат)

$$d_s = \frac{\sum_{j=1}^n |x_j - \bar{x}|}{n\bar{\sigma}} \quad (3)$$

Для завершення побудови показника, слід довести нерівність $d_s < 1$, Дана нерівність отримується завдяки нерівності Коші-Буняковського, $\left| \sum_{i=1}^n a_i b_i \right|^2 \leq \sum_{i=1}^n a_i^2 \sum_{i=1}^n b_i^2$. Покладаючи в ній $a_i = |x_i - \bar{x}|$, $b_i = 1$ отримуємо приведену нерівність. Отриманий показник (3) може бути використаний в створенні моделі показників, які функціонально залежать від середнього квадратичного відхилення, шляхом заміни змінних. Наприклад, із показника варіації ознаки $V = \frac{\sigma}{\bar{x}}$ можна, шляхом заміни x_i на $x_j - x$, враховуючи $x_i > 0$ і інваріантність виразу $x_j - x$, за такої заміни, ($x_j - \bar{x} - \overline{x_i - \bar{x}} = x_j - \bar{x} - \bar{x_i} + \bar{x} = x_j - \bar{x} - \bar{x} + \bar{x} = x_j - \bar{x}$), отримати формулу

$$V = \frac{\sigma}{\bar{x}} = \frac{\sqrt{\frac{\sum_{i=1}^n |x_i - \bar{x}|^2}{n}}}{\frac{\sum_{u=1}^n |x_i - \bar{x}|}{n}} = \frac{1}{d_s} \cdot \frac{\sum_{u=1}^n |x_i - \bar{x}|}{n\sigma} \dots$$

Стосовно прикладу, порівняння набору даних (заробітних плат), що належать до різних сукупностей, можна здійснювати за коефіцієнтом варіації даних, а також коефіцієнта варіації відхилень, з якого відбувається отримання коефіцієнта варіації самих даних.

Висновки та перспективи подальших досліджень. За статистичними даними в статті за модельним принципом відбулась побудова показника середньо квадратичного відхилення і на його основі коефіцієнта варіації. Аналогічно можна отримати походження і інших показників одновимірної описової статистики. Перспективи подальших

досліджень полягають у визначенні походжень показників багатовимірної дескриптивної статистики.

Список використаних джерел та літератури

1. Бахрушин В.Є Методи аналізу даних : навчальний посібник для студентів. – Запоріжжя: КПУ, 2011. – 268 с.2.
2. Іващенко П.О., Семеняк І.В., Іванов В.В. Багатовимірний статистичний аналіз. – Харків: Основа, 1992. – 144 с..