

В. М. Россочинський, старший викладач кафедри вищої математики та інформатики Академії митної служби України
О. М. Щитов, кандидат фізико-математичних наук, доцент кафедри інформаційних систем та технологій Академії митної служби України
Н. Г. Навроцька, кандидат технічних наук, доцент кафедри товарознавства та митної експертизи Академії митної служби України

ВИМІРЮВАННЯ ВАРІАЦІЇ АТРИБУТИВНОЇ ОЗНАКИ В МИТНІЙ СПРАВІ

Важливе значення для статистичних досліджень має однорідність сукупності, яка визначається показниками варіації. Для атрибутивних (якісних) ознак, що не мають безпосереднього кількісного вираження, вимірювання варіації актуально. У статті розглядаються властивості декількох характеристик, за допомогою яких можна вимірювати варіацію атрибутивних ознак.

The homogeneity of the set, which determined by the indices of variation has important meaning for statistical researches. For attributive characters do not have direct quantitative expression, the measuring of variation is actual. In the article there are considered the properties of a few characteristics of variation by which the measuring of variation of attributive characters is possible.

Ключові слова. Митна справа, статистичні дослідження, показники варіації, вимірювання варіації.

Вступ. Атрибутивною (якісною або номінальною) є ознака, “значення” якої не виражається числом, наприклад займана посада, спеціальне звання, стать, освіта, назва товару, що перетинає митний кордон, тощо. Той факт, що певна одиниця статистичної сукупності набуває певного “значення” атрибутивної ознаки, за якою дана сукупність досліджується, по суті означає, що дана одиниця сукупності має певну властивість, яка є одним із можливих “значень” ознаки.

Під час дослідження статистичних сукупностей за атрибутивною ознакою зазвичай будується атрибутивний ряд розподілу, тобто сукупність пар “властивість – частота” або “властивість – частка”. При цьому сукупність прийнято будувати у вигляді таблиці, в якій “значення” ознаки бажано розташовувати у певній логічній послідовності.

© **В. М. Россочинський, О. М. Щитов, Н. Г. Навроцька, 2009**

Частотою n_i i -го “значення” ознаки є кількість одиниць статистичної сукупності, що мають дане i -те “значення” (або дану властивість).

Частка (частість, або відносна частота) w_i i -го “значення” ознаки – це відношення частоти n_i до обсягу сукупності n . При цьому прийнято вважати, що “значення” трапляється в сукупності хоча б один раз, тобто частота $n_i > 0$ і, відповідно, частка (відносна частота)

$w_i > 0$ для всіх $i = \overline{1, m}$, де $m > 2$ – число різних “значень” ознаки. Якщо $n_k = 0$, то відповідне k -те “значення” ознаки виключається з атрибутивного ряду розподілу. За умови наявності лише одного “значення” ознаки, незалежно від її статистичної природи (якісна чи кількісна), варіація дорівнює 0.

Досліджуючи варіації кількісних ознак, тобто ознак, що мають безпосереднє кількісне вираження, у ролі числової міри варіації зазвичай застосовують певну систему числових характеристик, або показників [1–2], які, однак, мають певні недоліки, тому в деяких працях вводяться до розгляду нові показники варіації, що мають нові властивості [3].

Під час деяких статистичних досліджень виникає необхідність вимірювання величини варіації атрибутивної ознаки або її порівняння для декількох атрибутивних рядів розподілу. Наприклад, порівняння митних органів за рівнем освіти або за статтю його працівників.

В окремому випадку, коли атрибутивна ознака набуває лише двох взаємовиключних “значень” (тобто у випадку альтернативних ознак), в існуючій літературі як показник варіації розглядається дисперсія альтернативної ознаки [2, 4]. Крім того, певним аналогом показника варіації може виступати ентропія даної ознаки [2] як числова міра невизначеності даних спостереження.

Таким чином, завдання визначення показників для вимірювання варіації атрибутивної ознаки загалом досить актуальне.

Постановка завдання. Одиниці сукупності поряд із загальними для всіх одиниць ознаками мають індивідуальні особливості й відмінності, що відрізняють їх одна від одної. Відмінність значень досліджуваної ознаки одиниць сукупності називається варіацією даної ознаки. При цьому під варіацією атрибутивної ознаки розумітимемо наявність хоча б двох різних її “значень” з урахуванням їх часток w_i .

У процесі деяких статистичних досліджень виникає необхідність вимірювання величини варіації атрибутивної ознаки або порівняння цієї величини для різних атрибутивних рядів розподілу, що дає можливість судити про однорідність, у певному розумінні, досліджуваних сукупностей, сталість окремих значень ознаки, про взаємозв’язок між ознаками тощо.

Очевидно, що під час дослідження варіації атрибутивної ознаки необхідно мати числову міру (числову характеристику) для вимірювання величини варіації, яку надалі називатимемо характеристикою.

Узагалі характеристику варіації атрибутивної ознаки може бути введено неоднозначно, тобто можливе існування декількох різних видів характеристики, кожен з яких застосовуватиметься у певних межах (або у певних випадках) залежно від наявності (відсутності) у нього певних властивостей, які мають виражати характер залежності числового значення характеристики від величини варіації ознаки.

Мета даної роботи – введення у розгляд деяких нових характеристик варіації атрибутивної ознаки, дослідження їх властивостей та визначення можливостей застосування.

Результати дослідження. Проводячи дослідження з працівниками певного митного органу за атрибутивною ознакою “рівень освіти”, можна побудувати такий, наприклад, атрибутивний ряд розподілу (табл. 1).

Розподіл тих же працівників за атрибутивною ознакою “стать” може мати такий вигляд, як у табл. 2.

Таблиця 1

Розподіл працівників за рівнем освіти

№ з/п (i)	1	2	3	4	5	Разом
Рівень освіти	Середня	Молодший спеціаліст	Бакалавр	Спеціаліст	Магістр	×
Частота (n_i)	15	10	20	45	10	100
Частка (w_i)	0,15	0,10	0,20	0,45	0,10	1

Таблиця 2

Розподіл працівників за статтю

№ з/п (i)	1	2	Разом
Стать	Чоловіки	Жінки	×
Частота (n_i)	60	40	100
Частка (w_i)	3/5	2/5	1

Постає питання, як виміряти варіацію атрибутивної ознаки “рівень освіти”, що набуває 5 різних значень (див. табл. 1) або атрибутивної ознаки “стать”, яка має два значення (див. табл. 2).

Звичайно, бажано побудувати єдину універсальну характеристику, яка мала б загальноприйнятні властивості й була б застосовною в усіх можливих випадках. Але вибір такої характеристики ускладнюється тим, що перелік необхідних вимог до неї досить суб’єктивний і визначається двома видами чинників: 1) тими, від яких залежить величина варіації атрибутивної ознаки; 2) характером залежності значення характеристики від числових значень вищезгаданих факторів. Різні фахівці можуть мати різні думки щодо вищезазначених чинників, а це зумовлює можливість неоднозначного вибору характеристики.

На погляд авторів, логічно вважати, що величина варіації атрибутивної ознаки повинна збільшуватись: а) зі зростанням числа m різних її “значень”; б) зі зростанням рівноможливості цих “значень”, тобто, коли $w_i \rightarrow 1/m$ для всіх $i = \overline{1, m}$, досягаючи при фіксованому m найбільшої величини, коли $w_i = 1/m$.

Ураховуючи сказане, шукана характеристика повинна мати такі властивості (або задовольняти такі вимоги).

Властивість 1. Її числове значення має залежати тільки від двох факторів: числа m та співвідношення між частками w_i .

Властивість 2. Для фіксованого m характеристика має зростати, коли $w_i \rightarrow 1/m$, і досягати найбільшого значення при $w_i = 1/m$.

Властивість 3 може мати два різновиди:

3а) для рядів із різним числом m більше значення характеристики повинен мати ряд з більшим числом m незалежно від співвідношення між частками w_i ;

3б) зі збільшенням числа m значення характеристики має зростати. Але при цьому характер зростання може (або повинен) залежати також і від співвідношення між частками w_i , а саме:

- для двох рядів з різними значеннями числа m і рівноможливими “значеннями” ознаки більше значення характеристики повинен мати ряд із більшим числом m ;
- якщо “значення” ознак хоча б для одного ряду нерівномірні, то більше значення характеристики може мати ряд із меншим числом m , але більш рівномірними “значеннями” його ознаки.

Очевидно, що будь-яка характеристика повинна (або може) мати тільки один з різновидів властивості 3, оскільки вони взаємно несумісні. При цьому вибір залежить від власної думки дослідника щодо того, який із цих різновидів більш прийнятний у даному дослідженні (або взагалі).

Проте можуть існувати й інші погляди щодо вимог, які повинна задовольняти характеристика. У цій статті пропонуються до розгляду й можливого обговорення декілька конструкцій такої характеристики з аналізом їх основних властивостей.

1. Існує характеристика, яка має вигляд [2, 42]:

$$A = \prod_{i=1}^m w_i.$$

Слід зауважити, що окремим випадком величини A є вищезгадана дисперсія альтернативної ознаки.

Оскільки $w_i \in (0; 1]$, то при збільшенні числа m величина A загалом спадатиме. Але характер її зменшення залежатиме також і від співвідношення між частками w_i . Так, можлива ситуація, коли ряд із більшим числом m матиме більше значення A внаслідок більшої рівноможливості різних “значень” своєї ознаки. Цю властивість величини A можна розглядати як у певному розумінні протилежну до властивості 3б.

Для фіксованого m величина A збільшуватиметься, коли $w_i \rightarrow 1/m$, досягаючи найбільшого значення (див. примітку 1)

$$A_{\text{найб}} = m^m,$$

коли $w_i = 1/m$, тобто A має властивість 2.

Отже, величина A по-різному реагує на збільшення величини варіації ознаки за рахунок різних факторів: при збільшенні числа m вона загалом зменшується; при “вирівнюванні” значень w_i – збільшується. Очевидно, що наявність таких суперечливих властивостей робить величину A взагалі непринятною в ролі характеристики, але дозволяє використовувати її для порівняння величини варіації атрибутивних рядів з однаковим числом m (через наявність властивості 2).

II. На основі величини A можна побудувати характеристику

$$B = \frac{A}{A_{\text{найб}}} = m^{-m} \times \prod_{i=1}^m w_i.$$

З вигляду величини B впливають такі її властивості:

а) $B \in (0; 1]$;

б) для фіксованого m величина B зростає, коли $w_i \rightarrow 1/m$, і досягає свого найбільшого значення $B_{\text{найб}} = 1$ при $w_i = 1/m$ для всіх $i = \overline{1, m}$, тобто має властивість 2;

в) загалом не реагує на зміну величини варіації ознаки за рахунок зміни числа m , оскільки, по суті, являє собою частку, яку становить величина A від свого максимально можливого значення незалежно від m , тобто не має властивостей 1 і 3.

Останню властивість величини B можна оцінювати по-різному: позитивно, якщо вважати, що величина варіації повинна визначатись тільки співвідношенням між частками w_i і не залежати від m ; негативно – у протилежному разі.

У першому випадку величину B можна вважати цілком прийнятною характеристикою; в другому – використовувати тільки для порівняння варіації рядів з однаковим числом m .

III. Порівняно з попередньою більш вдалою можна вважати характеристику

$$C_k = m^k \times B = m^{m+k} \times \prod_{i=1}^m w_i.$$

де k – будь-яке додатне число.

З вигляду величини C_k очевидно, що зі зростанням m вона загалом збільшується й може набувати значення з множини $(0; m^k]$. Таким чином, C_k має властивості 1, 2 і різновид 3б властивості 3. І дійсно, значення C_k може бути більшим для ряду з більшим числом m , а може бути й меншим: оскільки хоча б одне зі значень w_i може бути як завгодно малим, то незалежно від m значення A , B та C_k також можуть бути як завгодно малими, що виключає наявність властивості 3а.

Отже, якщо вважати, що характеристика повинна мати властивості 1, 2 і 3б, то величина C_k може бути універсальною характеристикою для вимірювання величини варіації атрибутивної ознаки та її порівняння для будь-яких атрибутивних рядів розподілу.

IV. Розглянемо як характеристику величину

$$E = m + B = m^m \prod_{i=1}^m w_i.$$

Із властивостей величини B випливає, що величина E набирає значення з множини $(m; m + 1)$ і має властивості 1, 2, 3а.

Таким чином, якщо вважати, що характеристика повинна мати ці властивості, то величина E може використовуватись як універсальна характеристика для вимірювання варіації атрибутивної ознаки та її порівняння для будь-яких атрибутивних рядів розподілу.

Примітка 1.

Як відомо [5, 271], задача на умовний екстремум для функції m змінних

$$\Phi(x_1, x_2, \dots, x_m) = \prod_{i=1}^m x_i, \quad x_i > 0$$

за умови

$$\sum_{i=1}^m x_i = 1 \quad \text{або} \quad \varphi(x_1, x_2, \dots, x_m) = \sum_{i=1}^m x_i - 1 = 0$$

зводиться до задачі на безумовний екстремум для функції Лагранжа $(m + 1)$ змінної

$$L(x_1, x_2, \dots, x_m, \lambda) \equiv \Phi(x_1, x_2, \dots, x_m) + \lambda \times \varphi(x_1, x_2, \dots, x_m).$$

З необхідної умови екстремуму функції багатьох змінних одержуємо систему $(m + 1)$ алгебраїчних рівнянь для знаходження стаціонарних точок функції L :

$$\begin{cases} \frac{\partial L}{\partial x_k} = \prod_{i=1}^m x_i / x_k + \lambda = 0, \quad k = \overline{1, m}; \\ \frac{\partial L}{\partial \lambda} = \sum_{i=1}^m x_i - 1 = 0. \end{cases}$$

Із перших m рівнянь системи випливає, що всі x_i рівні між собою, а з останнього – що $x_i = 1/m$ для всіх $i = \overline{1, m}$.

Із виду функції Φ випливає, що вона не має мінімуму: $\Phi \rightarrow 0$, коли хоча б одна змінна $x_i \rightarrow 0$, але, оскільки $x_i > 0$ для всіх $i = \overline{1, m}$, то значення функції Φ може бути як завгодно близьким до нуля, не досягаючи його. Отже, стаціонарна точка $x_1 = x_2 = \dots = x_m = 1/m$ є точкою максимуму, і функція Φ досягає у цій точці найбільшого значення:

$$\Phi_{\text{макс}} = \prod_{i=1}^m \frac{1}{m} = m^{-m}.$$

Розглянемо визначення варіації групи імпортованих товарів, оформлених на митницях України за два періоди. При цьому візьмемо такі товари: чай товарної позиції 0902; розчинна кава товарної позиції 2101; різноманітні соуси товарної позиції 2103; фармацевтична продукція товарної позиції 3004; косметичні засоби товарної позиції 3304; засоби для прання, засоби для миття й чищення товарної позиції 3402. З метою зведення різних товарів до порівняльного вигляду у ролі “частоти” візьмемо вартість товару.

Дані для розрахунків [6, 7] наведено в табл. 3.

Таблиця 3

Розподіл імпортованих товарів, тис. дол. США

Код товару	0902	2101	2103	3004	3304	3402	Разом
Жовтень 2008	96 753	182 170	63 772	1 755 287	254 300	246 675	2 598 957
Листопад 2008	106 962	198 014	68 624	1 890 073	280 290	272 348	2 756 311

Вихідну інформацію перетворимо у вигляді часток w_i , тоді табл. 4 матиме вигляд:

Таблиця 4

Розподіл імпортованих товарів, тис. дол. США

Код товару	0902	2101	2103	3004	3304	3402	Разом
Жовтень 2008	0,037	0,070	0,025	0,675	0,098	0,095	1
Листопад 2008	0,039	0,072	0,025	0,663	0,102	0,099	1

Визначимо варіацію атрибутивної ознаки (назва товару) за допомогою встановлених характеристик, при цьому характеристикам, визначеним за жовтень, надамо позначку 1, а характеристикам, визначеним за листопад, – позначку 2.

$$A_1 = 0,037 \times 0,070 \times 0,025 \times 0,675 \times 0,098 \times 0,095 = 0,0000004069$$

$$A_2 = 0,039 \times 0,072 \times 0,025 \times 0,663 \times 0,102 \times 0,099 = 0,0000004698$$

$$B_1 = \frac{A_1}{A_{\text{жовтень}}} = m^m \times \prod_{i=1}^m w_i = 6^6 \times 0,0000004069 = 0,018984$$

$$B_2 = \frac{A_2}{A_{\text{листопад}}} = m^m \times \prod_{i=1}^m w_i = 6^6 \times 0,0000004698 = 0,021927$$

При $k = 1$:

$$C_{k1} = m^k \times B = m^{m+k} \times \prod_{i=1}^m w_i = 6^7 \times 0,0000004069 = 0,113906$$

$$C_{k2} = m^k \times B = m^{m+k} \times \prod_{i=1}^m w_i = 6^7 \times 0,0000004698 = 0,131564$$

$$E_1 = m + B_1 = m + m^m \times \prod_{i=1}^m w_i = 6 + 0,018984 = 6,018984$$

$$E_2 = m + B_2 = m + m^m \times \prod_{i=1}^m w_i = 6 + 0,021927 = 6,021927$$

Порівнявши обчислені характеристики за жовтень з відповідними характеристиками за листопад, бачимо, що всі вони мають меншу величину. Таким чином, за величинами обчислених характеристик обрана група товарів у листопаді має більшу варіацію, ніж у жовтні, тобто імпортованих товарів у листопаді більш різноманітний, ніж у жовтні.

Висновки. Якщо вважати, що характеристика повинна мати властивості 1, 2, 3а, (3б), то для вимірювання величини варіації атрибутивної ознаки, а також для її порівняння в різних атрибутивних рядах розподілу можна використовувати величину E (C_k).

Крім того, для порівняння величини варіації атрибутивної ознаки в рядах з однаковим числом різних її “значень” можна використовувати величину B .

Оскільки різні (навіть якщо вони подібні) характеристики можуть по-різному вимірювати й порівнювати величину варіації атрибутивної ознаки, то в кожному окремому дослідженні необхідно використовувати тільки одну попередньо вибрану характеристику.

Література

1. Ковтун Н. В. Загальна теорія статистики [Текст] : курс лекцій / Н. В. Ковтун, Г. С. Столяров. – К. : Четверта хвиля, 1996. – 144 с.
2. Теорія статистики [Текст] : учебник / под ред. Р. А. Шмойловой. – М. : Финансы и статистика, 2003. – 656 с.
3. Карамзин Н. Изучение вариации сублинейными отклонениями [Текст] / Н. Карамзин // Вестник статистики. – 1986. – № 6. – С. 63–65.
4. Статистика: теоретичні засади і прикладні аспекти [Текст] : навчальний посібник / за ред. Р. В. Фещура. – Львів : Інтеллект-Захід, 2003. – 576 с.
5. Кудрявцев Л. Д. Курс математического анализа [Текст] : учебник для студентов вузов / Л. Д. Кудрявцев : в 3 т. – М. : Высш. шк., 1988. – Т. 2. – 576 с.
6. Митна статистика зовнішньої торгівлі України [Текст] // Бюлетень. – 2008. – Жовтень. – 174 с.

7. Митна статистика зовнішньої торгівлі України [Текст] // Бюлетень. – 2008. – Листопад. – 176 с.