

**Key words:** *information, information resources, informative information, technologies, accounting, information accounting, information area.*

6

УДК 519.23:339.138

*П.М. Григорук, к.т.н., доцент, Хмельницький  
національний університет*

## **МЕТРИКИ ТА МІРИ ПОДІБНОСТІ ОБ'ЄКТІВ СПОСТЕРЕЖЕННЯ ПРИ ОПРАЦЮВАННІ МАРКЕТИНГОВИХ ДАНИХ**

*У роботі розглянутий аналіз основних підходів щодо визначення міри подібності об'єктів спостереження у багатомірному просторі ознак при розв'язанні завдань класифікації. Відзначені особливості визначення міри для даних нечислової природи.*

**Ключові слова:** *маркетингове дослідження, міра подібності, метрика, відстань в багатомірному просторі, класифікація.*

Маркетингові дослідження є засобом, що забезпечує менеджерів ринку інформацією, необхідною для ухвалення рішень. Вони грають ключову роль в реалізації маркетингового підходу до ведення бізнесу [1].

Загальна теорія та методологія маркетингових досліджень висвітлена у фундаментальних працях вітчизняних і зарубіжних вчених Г.П. Абрамової, О.Д. Андрєєвої, І.Ансоффа, Ф. Котлера, Є.В. Крикавського, Ж.Ж. Ламбена, Н.Малхотри, М.Портера, А.О. Старостіної, Г.Черчилля, Н.І. Чухрай.

Питання виявлення подібності об'єктів при розв'язанні завдань класифікації та групування розглянуті в роботах Б. Ван дер Вардена, І.Г. Венецького, Б.Г. Міркіна, М.Б. Мучника, А.І. Орлова, Ю.Н. Тюріна та інших.

**Метою даної статті** є розгляд найбільш вживаних метрик та міри подібності у багатомірному просторі ознак для розв'язання завдань класифікації об'єктів маркетингових спостережень.

В статистичних дослідженнях групування первинних даних є основним прийомом розв'язання завдань класифікації, а отже й основою проведення подальших досліджень. Завдання класифікації успішно розв'язуються за допомогою методів кластерного аналізу [2]. Він об'єднує сукупність методів, що дозволяють класифікувати багатовимірні спостереження за відсутності навчачих вибірок.

Напрямами застосування методів кластерного аналізу в маркетингових дослідженнях є [3]:

- сегментація ринку;
- групування торговельних марок і продуктів;
- позиціонування товару, що виводиться на ринок;

- вибір тестових ринків, на яких перевіряються висунуті припущення щодо реалізації маркетингових заходів;
- аналіз поведінки споживачів, об'єднаних в однорідні групи;
- розв'язання інших завдань, пов'язаних з необхідністю розбивати об'єкти на групи за відібраними в ході дослідження показниками.

Схожість об'єктів між собою у багатомірному просторі ознак залежить від їх геометричної близькості. Її можна оцінювати за мірою відстані між ними. В практиці маркетингових досліджень можна застосовувати досить велику кількість різноманітних шкал для вимірювання даних [4]. Якщо дані мають кількісну природу, їх доцільно виміряти в інтервальної шкалі. В табл. 1 наведені найбільш поширені міри відстані для такого випадку [2].

**Таблиця 1.**

**Міри відстані в багатомірному просторі для даних, виміряних за інтервальною шкалою**

Назва метрики	Формула для розрахунку
Евклідова відстань	$d_{ij} = \sqrt{\sum_{k=1}^n (z_{ik} - z_{jk})^2}$
Лінійна відстань	$d_{ij} = \sum_{k=1}^n  z_{ik} - z_{jk} $
Відстань Мінковського	$d_{ij}^{(p)} = \sqrt[p]{\sum_{k=1}^n  z_{ik} - z_{jk} ^p}$
Супремум-норма	$d_{ij} = \max_k  z_{ik} - z_{jk} $
Відстань Махаланобіса	$d_{ij}^{(M)} = (Z_i - Z_j)^T S^{-1} (Z_i - Z_j),$ де $S$ – коваріаційна матриця ознак
Відстань Брея-Кертіса	$d_{ij} = \sum_{k=1}^n  z_{ik} - z_{jk}  / \sum_{k=1}^n  z_{ik} + z_{jk} $
Канберрівська метрика	$d_{ij} = \sum_{k=1}^n \left(  z_{ik} - z_{jk}  /  z_{ik} + z_{jk}  \right)$

В цій і наступних таблицях під  $z_{st}$  будемо розуміти значення ознаки  $Z_s$  для об'єкта спостереження з номером  $t$ ,  $n$  – кількість ознак.

Найбільшого поширення набуло використання евклідової метрики. Разом з тим її застосування доцільне за умов, коли спостереження обираються з генеральної сукупності, що має багатомірний нормальний розподіл, а

ознаки  $Z_j$  взаємно незалежні і мають однакову дисперсію, а ознаки однорідні за змістом і однаково важливі для проведення аналізу.

Значення евклідової відстані істотно залежить від одиниць виміру ознак. Тому з метою вилучення впливу одиниць виміру вихідні ознаки початково стандартизують за формулою:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

де  $\bar{x}_j$  – середнє значення за  $j$ -тою вихідною ознакою,  
 $s_j$  – стандартне відхилення  $j$ -тої ознаки.

Відстань Махаланобіса використовується у випадках, коли необхідно врахувати взаємозв'язок між ознаками. Тобто, на відстань між будь-якими двома точками-об'єктами впливає не лише їх взаємне розташування в багатомірному просторі ознак, але й положення інших точок.

Лінійна відстань використовується у випадку, коли необхідно врахувати відмінності між об'єктами спостереження за кожною з ознак. Супремум-норма використовується тоді, коли виміряні об'єкти істотно розрізняються лише за однією з виміряних ознак. Відстань Мінковського є узагальненням наведених вище метрик. Як правило в практичних дослідженнях вона не використовується. Відстань Брея-Кертіса є модифікацією лінійної метрики і використовується тоді, коли необхідно звести відстані до діапазону  $[0; 1]$ . Відстань Канберра є нечутливою до істотних відхилень в діапазонах значень ознак для різних об'єктів. Це робить її особливо придатною для використання у випадку сильної асиметричності значень.

При проведенні маркетингових досліджень широкого вжитку набула дихотомічна шкала. Її використання є доцільним, коли оцінюється наявність або відсутність у об'єкта спостереження деякої властивості. Таким чином, дихотомічна ознака може приймати одне з двох значень: 0 або 1. Найбільш поширеною метрикою для дихотомічних даних є відстань Хеммінга, що розраховується за формулою

$$d_{ij} = \sum_{k=1}^n |z_{ik} - z_{jk}| \quad (2)$$

Вона використовується тоді, коли потрібно оцінити загальну кількість відмінностей за всіма координатами об'єктів.

Відстань між об'єктами в багатомірному просторі є мірою їх відмінності. чим більша відстань, тим більша відмінність між ними. Однак в дослідженнях часто потрібно оперувати поняттям подібності об'єктів. Перехід від відстані до подібності можна здійснити за формулою

$$\delta_{ij} = \frac{1}{1 + d_{ij}} \quad (3)$$

Отримані таким чином оцінки будуть знаходитись в межах від 0 до 1.

Для дихотомічних даних можна використовувати інші засоби оцінки подібності. Вони базуються на використанні чотириклітинної таблиці асоціативності, яка відображає спільну наявність або відсутність властивостей у об'єктів і має наступний вигляд (табл. 2).

**Таблиця 2.**

**Таблиця спряженості для двох об'єктів, виміряних за дихотомічною шкалою**

Значення змінних	1	0
1	a	b
0	c	d

Мір подібності, розрахованих за даними такої таблиці, існує досить велика кількість [5]. Найбільш вживані з них наведені в табл. 3.

**Таблиця 3.**

**Міри відстані в багатомірному просторі для даних, виміряних за дихотомічною шкалою**

Назва метрики	Формула для розрахунку
Міра Рассела-Рао	$d_{ij} = \frac{a}{a+b+c+d}$
Міра Хеммінга	$d_{ij} = \frac{a+d}{a+b+c+d}$
Міра Роджерса-Танімото	$d_{ij} = \frac{a}{a+b+c}$
Відстань Чекановського	$d_{ij} = \frac{b+c}{2a+b+c}$

Вибір конкретної метрики також залежить від завдання. Якщо важливим є виділення координат, для яких значення дихотомічних ознак співпадають, то перевагу слід надати мірі Хеммінга. Якщо перевага надається присутності властивості, то більш доцільним є використання мір Рассела-Рао або Роджерса-Танімото. Відстань Чекановського в чисельнику враховує відмінні одиничні значення за координатами, а в знаменнику – загальну кількість одиничних значень.

Для даних, виміряних в інших шкалах (номінальній, ранговій тощо) поняття відстані між об'єктами не має ніякого сенсу. Однак для них також

можна розрахувати показники подібності. Для розв'язання такого завдання можна також використати метрики, наведені в таблиці 4.

**Таблиця 4.**  
**Міри відстані в багатомірному просторі для даних, вимірних за дихотомічною шкалою**

Назва метрики	Формула для розрахунку
Відстань Журавльова	$d_{ij} = \sum_{k=1}^n I_{ij}^{(k)},$ де $I_{ij}^{(k)} = \begin{cases} 1, &  z_{ik} - z_{jk}  < \varepsilon \\ 0, & \text{інакше.} \end{cases}$
Відстань Вороніна	$d_{ij} = \sum_{k=1}^n \sigma_k \left  1 - \frac{ z_{ik} - z_{jk} }{(z_{k \max} - z_{k \min})} \right  / n,$ де $\sigma_k$ - вага k-тої ознаки

Метрика Журавльова має наступний сенс. Якщо об'єкти знаходяться досить близько за певною координатою, то відповідній частковій мірі ставиться у відповідність одиничне значення, інакше – нульове. Хоча в таблиці 4 розрахунок такої міри подібності поданий для кількісної шкали, її можна застосовувати і для інших шкал, де можна визначити ступінь близькості об'єктів за певною ознакою, наприклад, експертними методами.

Міра близькості Вороніна враховує вагові значення ознак, підкреслюючи їх нерівно значень при оцінці подібності. Фактично фона відображає різницю між координатами об'єктів відносно максимальної різниці доля відповідної ознаки. Таку міру також можна адаптувати для не кількісних ознак, керуючись наведеними вище міркуваннями.

Таким чином, в арсеналі дослідника маркетингової інформації є досить потужний арсенал різноманітних засобів для вимірювання подібності між об'єктами. Вибір конкретного інструментарію визначається наявними даними, обсягом вибірок та завданнями, які ставить перед собою дослідник.

#### Список використаних джерел

1. Малхотра Н.К. Маркетинговые исследования: практическое руководство, 4-е издание / Нереш К. Малхотра, Технологический институт Джорджии.– [Пер. с англ.] – [М.: ООО «И.Д. Вильямс», 2007]. – 1200 с.: ил. – [парал. тит. англ.].- ISBN 5-8459-0940-6 (рус).

2. Мандель И.Д. Кластерный анализ. / И.Д. Мандель. – М.: Финансы и статистика, 1988. – 176 с.: ил.– ISBN 5-279-00050-7.
3. Girish Punj. Cluster Analysis in Marketing Research: Review and Suggestions for Application / Girish Punj and David W. Stewart //Journal of Marketing Research, Vol. XX, (May 1983), pp.134-148.
4. Григорук П.М. Шкалювання в маркетингових дослідженнях. / П.М. Григорук // Вісник Хмельницького національного університету, 2009. – №5. – С. 138-142. – (Серія «Економічні науки», т.3).
5. А. Бююль. SPSS: искусство обработки информации. Platinum edition / Ахим Бююль, Петер Цёфель [Пер. с нем.]. – [СПб.: ООО «ДиаСофтЮП», 2005]. – 608 с.–ISBN 5-93772-133-0.

**Annotation.** *The analysis of basic approaches is in-process considered in relation to determination of supervision objects measure of similarity in multidimensional space of signs at the decision classification tasks. The features of measures determination are marked for information of non-numeric nature.*

**Key words:** *marketing research, measure of similarity, birth-certificate, multidimensional space distance, classification.*

7

УДК 336.22:334.012.6

*Л.Г. Гулько, к.е.н., доцент, Хмельницький національний університет*

## **МІЖНАРОДНИЙ ДОСВІД ПОДАТКОВОГО СТИМУЛЮВАННЯ МАЛОГО БІЗНЕСУ**

*У статті розглянуто роль малих підприємств в економіці України, фактори, що стримують його розвиток, а також міжнародний досвід податкового стимулювання малого бізнесу.*

**Ключові слова:** *малий бізнес, податки, стимулювання розвитку малого бізнесу.*

**Актуальність теми.** Малий бізнес відіграє вагомий роль в економічному розвитку країни. Проте в Україні розвиток підприємств малого бізнесу гальмується низкою перешкод, і зокрема фіскального характеру. Це зумовлює актуальність дослідження питань податкового стимулювання розвитку малого підприємництва.

**Аналіз попередніх досліджень і публікацій.** Дослідженню питань розвитку малого бізнесу та його оподаткування в Україні приділяють увагу такі вчені, як Варналій З. С., Говорушко Т. А., Кредісов В. А., Кужель О. В., Тимченко О. І. та ін.