УДК 519.23:339.138

ОЦІНКА ПОДІБНОСТІ ПОКАЗНИКІВ МАРКЕТИНГОВИХ ДОСЛІДЖЕНЬ

П.М. ГРИГОРУК

Хмельницький національний університет

У статті розглянуто найпоширеніші міри зв'язку для показників, отриманих в результаті маркетингових досліджень. Особлива увага надана ознакам, виміряним у нечислових шкалах. Отримані характеристики виступають мірами подібності між показниками для розв'язання різноманітних завдань статистичної обробки даних

Для будь-якої фірми, яка прагне до успіху, – маркетингові дослідження виступають як початок її діяльності. Маркетингові дослідження ринку значно зменшують невизначеність при ухваленні важливих маркетингових рішень, що дозволяє ефективно розподіляти економічний потенціал для досягнення нових висот в бізнесі. Вивчення зовнішнього і внутрішнього середовища і її регулярний моніторинг для будь-якого підприємства є важливим елементом стратегії успішного розвитку в умовах ринкової економіки. Роль маркетингових досліджень зростає у багато разів в умовах несформованості свого сегменту ринку або при невизначеності нового бізнесу.

Методології маркетингових досліджень присвячені численні праці вітчизняних і зарубіжних вчених Г.П. Абрамової, О.Д. Андрєєвої, І.Ансоффа. Г. Армстронга, Г.Л. Багієва, Ф. Котлера, Є.В. Крикавського, Ж.Ж.Ламбена, Н.Малхотри, М.Портера, А.О.Старостіної, Г. Черчилля, Н.І. Чухрай.

Маркетингові дослідження – це систематичне і об'єктивне виявлення, збір аналіз, розповсюдження і використання інформації для підвищення ефективності ідентифікації і вирішення маркетингових проблем і використання маркетингових можливостей [1]. Одним з його етапів є збір даних, який зазвичай здійснюється шляхом опитування. Його результатом є сукупність відповідей респондентів, яка по суті являє собою масив значень різноманітних показників, що характеризують середовище дослідження. Ця інформація є наслідком вимірювання відібраних характеристик. Тому процедура вибору адекватної шкали для отримання відповідного значення вимірюваному параметру є досить важливими для подальшого аналізу даних [2].

У більшості випадків дослідник має справу з характеристиками, які вимірюються за інтервальною або відносною шкалами. Такі дані зручні для подальшої статистичної обробки кількісними методами. Разом з тим невід'ємними характеристиками людей є різноманітні демографічні та соціально-економічні показники, такі як стать, сімейний стан, освіта, професія, посада, соціальний стан тощо. Зазвичай вони використовуються для перехресної класифікації даних з метою отримання якомого повнішої інформації з відповідей респондента. Одні з них (освіта, професія) можуть бути встановлені точно і досить легко, інші – лише наближено (наприклад, соціальний статус особи). Визначення кількісних характеристик таких показників для їх статистичної обробки є досить непростим завданням. При цьому важливо виявити взаємозв'язки між ними, їх вплив на інші показники або поведінку респондентів тощо.

Метою цієї статті є розгляд мір подібності ознак у багатомірному просторі для розв'язання завдань стиснення вихідного масиву ознак без істотної втрати їх інформативності, пошуку латентних ознак, виявлення взаємозв'язку між ознаками з метою побудови різноманітних залежностей, знаходження найбільш впливових та репрезентативних показників тощо.

Для об'єктів спостереження подібність можна оцінювати за мірою відстані між ними у багатомірному просторі. Для ознак поняття відстані між ними не має практичного смислу. Тому подібність можна оцінювати за значеннями коефіцієнта кореляції. Він є мірилом статистичного взаємозв'язку між декількома випадковими величинами, в ролі яких виступають досліджувані показники.

Якщо ознаки виміряні за інтервальною або відносною шкалами, то використовується коефіцієнт парної кореляції Пірсона [3]. Однак його можна застосовувати лише тоді, якщо є відомості, що форма зв'язку між ознаками – лінійна. Крім того, при розрахунку цього коефіцієнта побічно враховується взаємозв'язок показників з іншими ознаками, що є певним недоліком. Якщо цей вплив потрібно вилучити, то використовуються часткові коефіцієнти кореляції, які розраховуються за формулою

$$r_{ij,k} = \frac{r_{ij} - r_{ik} r_{jk}}{\sqrt{1 - r_{ik}^2}\sqrt{1 - r_{jk}^2}} ,$$

де $r_{ij,k}$ – коефіцієнт часткової кореляції між показниками X_i та X_j за умови вилучення впливу X_k;

r_{st} – коефіцієнт парної кореляції між показниками X_s та X_t .

В деяких випадках потрібно вилучити вплив іншого показника лише з одної зі змінних, для яких розраховується показник взаємозв'язку. В таких випадках доцільно використовувати частинний коефіцієнт кореляції. Якщо для змінних X_i та X_j потрібно вилучити вплив X_k лише зі змінної X_i ,то відповідний показник розраховується за формулою

$$r_{j(i,k)} = \frac{r_{ij} - r_{ik} r_{jk}}{\sqrt{1 - r_{ik}^2}} .$$

Якщо досліджується взаємозв'язок одного з показників з деякою кількістю інших, то використовується коефіцієнт множинної кореляції. Наприклад, для обчислення сили зв'язку змінної X_i зі змінними X_j та X_k він матиме вигляд

$$r_{i,jk} = \sqrt{\frac{r_{ij}^2 + r_{ik}^2 - 2 r_{ij} r_{ik} r_{jk}}{1 - r_{jk}^2}} .$$

Аналогічно можна оцінити множинний зв'язок між більшою кількістю показників, але відповідні залежності будуть мати досить складний вигляд, тому в дані статті їх не розглядатимемо.

Як відзначалось раніше, маркетингові дані можуть мати нечислову природу. Якщо вони виміряні за порядковою шкалою, то для вивчення взаємозв'язку між ними використовуються показники неметричної, або рангової кореляції. В таблиці 1 наведені найбільш вживані з таких мір [3]. Для їх розрахунку розраховуються ранги R_i та R_i^* показників X_i та X_j відповідно.

Найбільш розповсюдженими є коефіцієнти рангової кореляції Спірмена та Кендалла. Вони не вимагають складних розрахунків і є інтуїтивно зрозумілими. Коефіцієнти конкордації Кендалла-Сміта та Шукені-Фролі виступають в ролі показників множинної кореляції, коли вимірюється зв'язок між групою показників. Остання характеристика часто використовується в експертному оцінюванні для узгодження думок експертів, коли дві групи експертів чисельністю n_1 та n_2 оцінюють один і той же показник.

Таблиця 1. **Показники рангової кореляції**

Назва коефіцієнта	Формула для розрахунку
Коефіцієнт Спірмена	$\rho = 1 - \frac{6\sum_{j=1}^{m} d_j^2}{m(m^2-1)}$, де $d_j = R_j - R_j^*$
Коефіцієнт Кендалла	$\tau = \frac{4\sum_{j=1}^{m-1} S_j}{m(m-1)} - 1$, де S_j – кількість рангів другого показника, що знаходяться справа від значення X_{2j} і більших за R_j^*; припускається, що при цьому показники впорядковані за зростанням рангів першого показника X_1
Коефіцієнт конкордації Кендалла-Сміта	$W = \frac{12S_W}{k^2(m^3-m)}$, де $S_W = \sum_{i=1}^{m}\left(\sum_{j=1}^{k} R_{ji} - \frac{k(m+1)}{2}\right)^2$, k – кількість показників, для яких вимірюється міра зв'язку, R_{ji} – значення рангу i-того показника на j-тому об'єкті спостереження
Коефіцієнт конкордації Шукені-Фролі	$W = \frac{L - M(L)}{L_{\max} - M(L)}$, де $L = \sum_{i=1}^{m} R_i R_i^*$, $R_i = \sum_{k=1}^{n_1} R_{ki}$, $R_i^* = \sum_{k=1}^{n_2} R_{ki}^*$, $L_{\max} = \frac{n_1 n_2 (m+1)(2m+1)}{6}$, $M(L) = \frac{n_1 n_2 m(m+1)^2}{4}$

Якщо показники виміряні за дихотомічною шкалою, міру зв'язку між ними можна оцінити за таблицею асоціативності, яка має вигляд, наведений в таблиці 2. Вона відображає спільну наявність або відсутність властивостей ознак у об'єктів спостереження. Основні міри подібності, розраховані за таблицею асоціативності, наведені в таблиці 3.

Таблиця 2. **Таблиця асоціативності для двох об'єктів, виміряних за дихотомічною шкалою**

Значення змінних	1	0
1	a	b
0	c	d

Таблиця 3. **Показники близькості ознак, розраховані за таблицею асоціативності**

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

Вибір конкретної метрики також залежить від завдання. Якщо важливим є врахування значень,, які співпадають, то перевагу слід надати мірі Хеммінга. Коли ж перевага надається одиничним значенням, 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.

Наведені показники можна використовувати і для даних, виміряних за ранговою або номінальною шкалами, якщо вони містять два значення.

Якщо необхідно оцінити зв'язок між ознаками, які мають більшу , ніж дві, кількість градацій за кожною з них, використовуються багатоклітинні таблиці асоціативності. Нехай така таблиця має p рядків та q стовпчиків. Позначимо через n_{ij} кількість спостережень, що мають i-ту та j-ту градації властивостей одночасно. Тоді загальна міра зв'язку оцінюється за коефіцієнтом квадратичної зв'язаності χ^2, який розраховується за формулою

$$\chi^2 = n\left(\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{n_{ij}}{n_i n_j^*} - 1\right),$$

де n_i – сума значень таблиці за i-тим рядком; n_j^* сума значень таблиці за j-тим стовпчиком, n – загальна кількість значень

Зазвичай коефіцієнт квадратичної зв'язаності використовується як допоміжний для розрахунку інших показників зв'язку [1], основні з яких наведені в таблиці 4. Найбільшого поширення серед них набули коефіцієнти Крамера та Чупрова.

Якщо необхідно виміряти зв'язок між ознаками, що виміряні в різних шкалах, потрібно скористатись спеціальними показниками. Для кореляції змінних, одна з яких виміряна у інтервальні шкалі, а інша – у дихотомічній, використовують точково-бісеріальний коефіцієнт кореляції. Він розраховується за формулою

$$r_{pb} = \frac{\overline{X}_1 - \overline{X}_0}{s_X}\sqrt{\frac{m_1 m_0}{m(m-1)}} ,$$

де $\overline{X}_1, \overline{X}_0$ – середні значення ознаки, виміряної за інтервальної шкалою зі значеннями 1 та 0 іншої ознаки відповідно;

s_X – стандартне відхилення всіх значень ознаки, виміряної за інтервальної шкалою;

m_1, m_0 – кількість значень першої ознаки зі значеннями 1 або 0 дихотомічної ознаки відповідно;

m – загальна кількість пар значень.

Таблиця 4. **Показники близькості ознак, розраховані за багатоклітинною таблицею асоціативності**

Назва показника	Формула для розрахунку
Коефіцієнт φ	$\varphi = \sqrt{\chi^2 / n}$
Коефіцієнт зв'язаності Пірсона	$C = \sqrt{\chi^2 /(\chi^2 + n)}$
Коефіцієнт Крамера	$V = \sqrt{\dfrac{\chi^2}{n \cdot \min(p-1, q-1)}}$
Коефіцієнт Чупрова	$T = \dfrac{\chi^2}{n\sqrt{(p-1)(q-1)}}$

Випадки, коли одна із змінних представлена в ранговій шкалі, а інша в дихотомічній, вимагають застосування коефіцієнта рангово-бісеріальної кореляції**,** який розраховується за формулою

$$r_{pb} = \frac{2(r_1 - r_0)}{m},$$

де m – кількість об'єктів вимірювання;

r_1, r_0 – середні значення рангів першої ознаки зі значеннями 1 або 0 за другою ознакою відповідно.

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

ЛІТЕРАТУРА

1. Малхотра Н.К. Маркетинговые исследования: практическое руководство, 4-е издание / Нереш К. Малхотра, Технологический інститут Джорджии.– [Пер.с англ.]– [М.: ООО «И.Д. Вильямс», 2007].– 1200 с.: ил.– [парал. тит. англ].- ISBN 5-8459-0940-6(рус).
2. Григорук П.М. Шкалювання в маркетингових дослідженнях / П.М. Григорук // Вісник Хмельницького національного університету, 2009. – .№5.-с.138-142. – (Серія «Економічні науки», т.3).
3. Кобзар А. И. Прикладная математическая статистика. Для инженеров и научных работников / А.И. Кобзарь.– М.:ФИЗМАТЛИТ, 2006.– 816 с.– ISBN 5-9221-0707-0.

Надійшла 21.10.2010