

**Н.О. Тіхонова,**

ас. каф. менеджменту НУХТ

**N. Tichonova**

Assistant of the cathedra of management

*juzik\_n@mail.ru*

**ВИКОРИСТАННЯ КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ПРОЕКТУВАННЯ  
ПРОДУКТОВИХ ПОРТФЕЛІВ М'ЯСОПЕРЕРОБНИХ ПІДПРИЄМСТВ  
USING OF CLUSTER ANALYS FOR PROJECTING PRODUCT  
PORTFOLIOS OF MEAT PROCESSING FACTORIES**

*В статті розглядаються проблеми:*

- *проектування продуктових портфелів підприємств м'ясопереробної галузі;*
- *визначення з максимальним ступенем точності істотних факторів, що впливають на виробництво ковбасних виробів;*
- *складання науково обґрунтованого прогнозу динаміки «поведінки» складових продуктового портфелю м'ясопереробних підприємств, базуючись на вивченні цих факторів;*
- *формування оптимального продуктового портфеля, який дозволяє враховувати сучасні тенденції в сфері класифікації продуктів харчування.*

*Для вирішення цих питань пропонується застосування кластерного аналізу, як найбільш вдалого методу побудови класифікацій та оптимізації структури продуктових портфелів підприємств. Враховуючи особливості та переваги цього методу представляється перспективним його застосування для моделювання та оптимізації продуктового портфелю саме м'ясопереробних*

*підприємств.*

*Ключові слова: продуктовий портфель, кластерний аналіз, однорідні групи, концепція здорового харчування, фактори впливу.*

*In the article there are opened such problems:*

- projecting product portfolios of meat processing factories;*
- figuring out, with highest level of accuracy, main factors, which have influence on manufacturing sausage products;*
- making substantiate forecast of the dynamic of conduction of components in product portfolios of meat processing factories, based on discovering this factors;*
- forming optimally product portfolio, which helps to use modern ways in the sphere of classification food products.*

*For making this problems open we propose to use cluster analyze, as the best one for making classifications and structure optimizations product portfolios of factories. Because of advantages of this method we can be convinced of its usefulness in the future for modeling and optimizing product portfolios exactly in meat processing factories.*

*Key words: product portfolio, cluster analyze, similar groups, conception of healthy nutrition, influence factors.*

**Вступ.** Асортимент, структура і склад продуктів харчування, в тому числі м'ясних виробів, значно змінюються у відповідності з вимогами науки про харчування, положеннями концепції здорового харчування та економічного стану суспільства. Не дивлячись на формальну наявність стандартів, які регламентують різні характеристики сировини, допоміжних матеріалів і готових виробів, їх якість у більшості випадків не відповідає світовому рівню, а структура асортименту м'ясопереробних підприємств - вимогам споживача щодо якості харчових продуктів, їх смаку, аромату, текстурі, відповідності національним традиціям.

При цьому, як свідчить закордонний досвід, здатність м'ясопереробних підприємств до виживання в умовах конкуренції,

економічна ефективність їх діяльності багато в чому визначається якістю та оптимальністю структури продукції, що виробляється. Створення м'ясних виробів, які відповідають сучасним вимогам, можливе тільки при раціональному використанні сировини, інгредієнтів, врахуванні сучасної концепції раціонального харчування, що відображає добову потребу в нутрієнтах. У зв'язку з цим виникає потреба у проектуванні продуктового портфелю підприємств галузі.

**Постановка завдання.** Методичним підґрунтям для проектування продуктового портфелю вітчизняних м'ясопереробних підприємств може стати кластерний аналіз, тому що зміна структури в результаті впливу різних тенденцій може виявитись достатньо складною з великою ймовірністю та погано передбачуваною, якщо використовувати звичайний кореляційно-регресійний аналіз.

**Результати.** Як теоретики, так і практики, що займаються оптимізацією структури виробництва, регулярно стикаються з труднощами, коли перед ними виникає задача розподілу множини існуючих асортиментних позицій на різні групи з відносно однорідною структурою та зміною її у часі. Основною проблемою є питання підбору та узгодження відібраних факторів так, щоб їх представлення в багатомірній системі координат достатньо точно виконувало розбивку на кластери, які характеризуються максимально подібними тенденціями. При цьому потрібно враховувати, що навіть за умови вдалого підбору точних коефіцієнтів для існуючих кількісних факторів, завжди знайдуться не менш важливі якісні показники, представити які в кількісному вигляді практично неможливо. Розподіл множини асортиментних позицій на окремі кластери в залежності від динаміки структури здійснюється таким чином: дані по структурі асортименту за досліджувані періоди представляються у вигляді матриці:

$$\begin{pmatrix} R_{10} & R_{11} & \dots & R_{1r} \\ R_{20} & R_{21} & \dots & R_{2r} \\ \dots & \dots & R_{km} & \dots \\ R_{N0} & R_{N1} & \dots & R_{Nr} \end{pmatrix} \quad (1)$$

де  $R_{km}$  – частка  $k$ -ї асортиментної позиції за  $m$ -ї період,  $k \in [1; N]$ ;  $m \in [0; T]$

Далі, формування кластерів відбувається через вираховування євклідової відстані між асортиментними позиціями  $p$  та  $q$  за формулою:

$$r_{p,q} = \sqrt{\sum_{m=1}^T \left( \frac{R_{pm} - R_{qm}}{\sigma_{Rm}} \right)^2} \quad (2)$$

де  $m$  – номер періоду,  $\sigma_{Rm}$  – середньоквадратичне відхилення частки за період  $m$ . Критична величина розбивки передбачається рівною квадратному кореню з кількості періодів  $T$ , тобто середній величині євклідової відстані [1]:

$$r_{крит} = \overline{r_{p,q}} = \sqrt{T} \quad (3)$$

Переваги даної методики полягають, по-перше, в тому, що вона дозволяє з дуже високим ступенем точності групувати асортиментні позиції зі схожими тенденціями в зміні структури протягом всього періоду, який визначає базу прогнозу, що надає підстави розраховувати на збереження подібної тенденції і в майбутньому. Другою її перевагою є можливість повної автоматизації, що значно полегшує роботу, дозволяючи використовувати сучасні обчислювальні засоби, а також обробляти однорідну інформацію, отримвану з електронних баз даних. Тому вона може бути впроваджена без значних труднощів не тільки в комп'ютерних системах окремих підприємств м'ясопереробної галузі, які виробляють значний асортимент продукції, але і на інших підприємствах харчової промисловості, де можуть виникнути проблеми із формуванням структури продуктового портфелю. Напевно, найбільш гострою проблемою, яка виникає перед спеціалістами по факторному аналізу, є

підбір чітких критеріїв, які дозволяють позбавитись від другорядних факторів, що збільшують розмірність моделі без зростання її точності, і при цьому правильно визначити вагу для інших факторів. Доказом важливості цього питання, а також відсутності однозначно оптимальних рішень, є множина існуючих критеріїв відбору значущих компонент. Достатньо назвати такі відомі методи, як розрахунок варімакс-критерію,  $n$ -критерій, відбір за  $t$ -критерієм Стюдента тощо. Таким чином, використовувати в моделі черговий фактор доцільно тільки в тому випадку, якщо він в достатньому ступені знижує рівень ентропії, отже, збільшується значення  $R$ -квадрат. Яким чином, чисельно визначити приріст даної величини в залежності від кількості використаних факторів? Розглянемо цю проблему в світлі коефіцієнтів послідовної детермінації. Нехай, маємо  $N$  факторів  $X_1, \dots, X_N$ , які ймовірно впливають на доходність підприємства при існуючій структурі продуктового портфеля (умовно будемо рахувати, що цей же показник і є показником доходності такого портфеля). При введенні в рівняння регресії фактора  $X_i$  показник  $R$ -квадрат приймає деяке значення. Оберемо фактор, при якому він буде найбільшим [1]:

$$p_1^2 = R_1^2 = r_{yx_1}^2 \quad (4)$$

де  $p_1^2$  - коефіцієнт послідовної детермінації для даного фактора,  $r_{yx_1}^2$  - парний коефіцієнт кореляції між доходністю портфелю та цим фактором. Тепер введемо в отримане рівняння регресії другий фактор таким чином, щоб значення  $R$ -квадрат знову стало максимально можливим потім розрахуємо другий коефіцієнт послідовної детермінації:

$$p_2^2 = R_2^2 - R_1^2 = r_{yx_1x_2}^2 - r_{yx_1}^2 \quad (5)$$

Аналогічно розраховуються наступні коефіцієнти:

$$p_3^2 = R_3^2 - R_2^2 = r_{yx_1x_2x_3}^2 - r_{yx_1x_2}^2 \quad (6)$$

$$p_4^2 = R_4^2 - R_3^2 = r_{yx_1x_2x_3x_4}^2 - r_{yx_1x_2x_3}^2 \quad (7)$$

тощо [1]. Базовий відбір факторів продовжується доти, поки величина отриманих коефіцієнтів послідовної детермінації не стане меншою деякого критичного значення. Враховуючи, що в механізм розрахунку скорегованої величини *R-квадрат* входить поправка на збільшення ентропії при додаванні нових факторів, її приріст на кожній ітерації алгоритма повинен бути позитивним і, як наслідок, критичне значення *p* повинно бути більше нуля. Даний метод дозволяє відібрати зі всіх факторів саме ті, які найбільше впливають на доходність досліджуваної структури продуктового портфелю. Це дозволяє істотно понизити розмірність моделі, створюваної на основі методики, прискорити розрахунки і при цьому відкинути дані, які не мають великого впливу на досліджувані показники. Як правило, від виявлених головних компонент залежить не менше 85% загальної дисперсії, що доводить ефективність обраного методу аналізу. Тепер, коли визначені методи відбору факторів і технологія розшарування множини асортиментних позицій продуктового портфелю на окремі кластери, можна розробити методичні підходи до оптимізації асортименту продуктового портфелю. Враховуючи, що в теперешній час впровадження будь-якої економічної методики неможливо без автоматизації, існує алгоритм, за яким належить виконувати операції для отримання шуканого результату: оптимізованої структури продуктового портфелю, яка дозволяє отримати максимальний прибуток при заданому рівні ризику. На першому етапі визначимо масиви даних, які підлягають математичній обробці. Отже, маємо такі вихідні дані:  $S_1, S_2, \dots, S_N$  – досліджувана множина асортиментних позицій;

$$\begin{pmatrix} R_{10} & R_{11} & \dots & R_{1T} \\ R_{20} & R_{21} & \dots & R_{2T} \\ \dots & \dots & R_{km} & \dots \\ R_{N0} & R_{N1} & \dots & R_{NT} \end{pmatrix} \quad (8)$$

матриця часток асортиментних позицій  $S_1-S_N$  за періоди  $[0 ; T]$ ,

де  $R_{ij}$  – частка  $i$  за  $j$ -й період;

$$\begin{pmatrix} X_{10} & X_{11} & \dots & X_{1T} \\ X_{20} & X_{21} & \dots & X_{2T} \\ \dots & \dots & X_{ij} & \dots \\ X_{K0} & X_{K1} & \dots & X_{KT} \end{pmatrix} \quad (9)$$

матриця факторів  $X_1-X_K$  за періоди  $[0 ; T]$ ,

де  $X_{ij}$  – значення фактора  $X_i$  за  $j$ -й період [1];

Тепер необхідно визначити частки  $m_1, \dots, m_N$  існуючих в продуктовому портфелі асортиментних позицій з метою максимізації доходності в наступному періоді:

$$\begin{cases} \sum_{i=1}^N m_i \cdot ER_i \rightarrow \max \\ E\sigma_\pi = \sigma_\pi^* \end{cases} \quad (10)$$

де рівень доходності  $R_i$  вираховується як відношення очікуваної в звітному періоді фактичної ціни виду продукції  $S_i$  до попередньої (або запланованої) в момент формування портфелю за вирахуванням одиниці. Так, доходність за місяць в момент часу  $t=1$  вираховується таким чином [1]:

$$R_{i1} = \frac{Pr_{i1}}{Pr_{i0}} - 1 \quad (11)$$

Розробимо алгоритм оптимізації продуктового портфеля з застосуванням кластерного аналізу. Такий алгоритм можна умовно розподілити на чотири стадії:

- 1) Розшарування множини асортиментних позицій на окремі кластери;
- 2) Визначення факторів, які впливають на склад кожного кластеру;
- 3) Прогнозування динаміки відібраних факторів;
- 4) Визначення найбільш перспективних груп на основі співставлення дендрограм для сукупності проаналізованих підприємств;

5) Визначення оптимального набору асортиментних позицій та їх частки в продуктовому портфелі для забезпечення максимізації доходності.

Тепер розглянемо ці стадії детальніше:

1. Стадія розшарування множини асортиментних позицій на окремі кластери починається з формування таблиці євклідових відстаней між ними.

Дві асортиментні позиції з найменшими відстанями об'єднуються в кластер, і середню частку якого і середню доходність можна вирахувати за формулою середньої арифметичної, після чого процедура повторюється. Процес об'єднання в кластери припиняється, коли мінімальна відстань між групами перебільшить критичне значення [2]:

$$r_{кр} = \sqrt{T} \quad (12)$$

В результаті такої процедури, замість «випадкової» множини асортиментних позицій, отримуємо набір впорядкованих кластерів, об'єднаних на основі загальних тенденцій. При цьому досягається відразу дві важливі цілі: по-перше, значно скорочується кількість змінних, що в значній мірі спрощує розрахунки, а по-друге, зменшується частка впливу випадкових факторів, які можуть в окремі момент часу корелювати з часткою та доходністю окремих видів продукції. В межах кластеру за рахунок виконаної диверсифікації ймовірність випадкових збігів зменшується в багато разів, що надає можливість більш чітко визначити фактори, які реально впливають на структуру та доходність.

- 2) Для того, щоб обрахувати величину впливу кожного фактора на відповідий кластер, представимо доходність по кластерам в наступному вигляді:

$$R'_t = F_1 \cdot X_{1t} + F_2 \cdot X_{2t} + \dots + F_k \cdot X_{kt} + \varepsilon_t \quad (13)$$

де  $F_i$  – коефіцієнт фактора  $X_i$  в рівнянні множинної регресії,  $\varepsilon_t$  – помилка в період часу  $t$  [2]. При цьому величина  $T$  повинна значно (не менше ніж в п'ять разів) перевищувати кількість факторів  $k$ . Значущі фактори

відбираються за допомогою методу із застосуванням коефіцієнтів послідовної детермінації. Фактори відбираються послідовно, а вибір визначається шляхом максимізації коефіцієнта:

$$p_i^2 = R_i^2 - R_{i-1}^2 = r_{yx_1x_2\dots x_i}^2 - r_{yx_1x_2\dots x_{i-1}}^2 \quad (14)$$

Процес додавання факторів продовжується до того моменту, доки максимальний скорегований коефіцієнт послідовної детермінації не стане від'ємною величиною. Для будь-якої відібраної кількості факторів коефіцієнти  $F_1, F_2, \dots, F_k$  розраховуються таким чином, щоб мінімізувати суму квадратів помилок регресії за період бази прогнозу [2]:

$$\sum_{t=0}^T \varepsilon_t^2 \rightarrow \min \quad (15)$$

Цієї мети можна досягти шляхом математичних перетворень матриці факторних ваг. На сьогоднішній день існує низка програмних пакетів, які дозволяють виконувати такі розрахунки з високою швидкістю і за короткий час. Результатом реалізації цього алгоритму є отримання формул множинної регресії для кожного кластера, з допомогою яких, спираючись на статистичні дані про динаміку факторів, можна отримати прогноз розвитку доходності кластерів на наступний період.

4. Четвертий етап буде присвячений переходу від вивчення загальних кластерних тенденцій до розрахунку індивідуальних рівнянь регресії для кожної асортиментної позиції. Розрахунок очікуваної доходності і ступеня ризику для кожної асортиментної позиції. При цьому розрахуємо кореляцію доходності із середньою доходністю по кластеру. В даних розрахунках доцільно застосувати бета-коефіцієнт:

$$\beta_i = r_{ic} \cdot \frac{\sigma_i}{\sigma_c} \quad (16)$$

де  $r_{ic}$  – коефіцієнт кореляції між доходністю окремого виду продукції та середньою доходністю кластера, до якого вона належить,  $\sigma_i$  та  $\sigma_c$  – відповідно їх середньоквадратичні відхилення [3].

Після розрахунку бета-коефіцієнта доходність кожного із досліджуваних видів продукції можна буде передбачити за допомогою такого рівняння регресії:

$$R_{it} = \alpha_i + \beta_i \cdot \check{R}_{ct} + \varepsilon_{it} \quad (17)$$

а очікувана в наступному періоді доходність буде визначатись за формулою:

$$E_T R_{it+1} = \alpha_i + \beta_i \cdot E_T \check{R}_{ct+1} \quad (18)$$

При цьому коефіцієнт невизначеності для досліджуваної продукції:

$$k_{nd} = 1 - r_{ic}^2 \quad (19)$$

а величина ризику –

$$\sigma_i = \sqrt{\frac{\sum_{t=1}^T (E_T R_{it} - R_{it})^2}{T}} \quad (20)$$

5. Після всіх проведених розрахунків отримуємо для кожної асортиментної позиції величину очікуваної доходності. Тепер задачу можна звести до визначення частки асортиментної позиції в продуктовому портфелі з метою максимізації доходності:

$$\sum_{n=1}^N (E_T R_{iT+1}) \rightarrow \max \quad (21)$$

$$\sqrt{\sum_{i=1}^N \sum_{j=1}^N m_i m_j V_{ij}} = \sigma_\pi \quad (22)$$

де  $m_i$  - частки  $\sum_{i=1}^N m_i = 1$ .

**Висновки.** Кластерний аналіз включає в себе набір різних алгоритмів класифікації. Загальне питання, яке задають дослідники різних областей полягає в тому, як організувати дані спостережень в наглядні структури. Кластерний аналіз необхідний для класифікації інформації, з його допомогою можна структурувати змінні і дізнатись, які з них поєднуються в першу чергу, а які слід розглядати окремо. Велика перевага цього методу полягає в тому, що він дозволяє провести групування не за однією ознакою, а за набором різних ознак.

Запропонована методика дозволяє оптимально розв'язати відразу дві важливіші проблеми: розшарування множини асортиментних позицій на окремі однорідні групи, а також виявлення факторів впливу зовнішнього середовища, які впливають на групи з послідовним знаходженням факторних ваг. Це дозволяє уникнути штучної дискретності, яка виникає при жорсткому відборі факторів зовнішнього середовища та сортуванні видів продукції виключно за асортиментними агрегатами. Технологія портфельної оптимізації дозволяє формувати продуктивний портфель на основі обраного ступеня ризику та прогнозів змін основних факторів. Не менший практичний інтерес представляє порівняння існуючого продуктового портфеля з еталонним портфелем в розрізі обраних факторів, що дозволяє виявити та оцінити переваги і недоліки досліджуваного асортименту, користуючись при цьому кількісними критеріями.

#### **ЛІТЕРАТУРА**

1. Олдендерфер М.С., Блэшфилд Р.К. Кластерный анализ / Факторный, дискриминантный и кластерный анализ: пер. с англ.; Под. ред. И.С. Енюкова. - М.: Финансы и статистика, 2002 — 215 с.
2. Таганов Дмитрий . SPSS: статистический анализ в маркетинговых исследованиях.- П., 2005. – 192 с.
3. Шморгун Н.П., Головка В.І. Фінансовий аналіз: Навч. посіб. – К.: Центр навчальної літератури, 2006. – 450 с.

*Надійшла до редколегії 25.06.2012*

