

УДК.658:001.89

Мельниченко О.П., кандидат с.-г. наук, доцент кафедри вищої математики
(e-mail:mela731@rambler.ru)

Білоцерківський національний аграрний університет

СУТНІСТЬ КОРЕЛЯЦІЙНО-РЕГРЕСИВНОГО МЕТОДУ СТАТИСТИЧНИХ ДОСЛІДЖЕНЬ

Представлена робота є одним із можливих варіантів застосування методів математичної статистики та створення математичних моделей у різних галузях науки. У ній детально пояснено основні етапи дослідницької діяльності, які можуть бути застосовані до сільського господарства та промисловості.

Ключові слова: *кореляційно-регресійний аналіз, математична модель, рівняння ліній регресії, фізико-хімічні параметри молока.*

Навчання з математичної статистики є важливою частиною методології підготовки наукових кадрів у питаннях якісно-кількісного аналізу масових явищ. Оволодіння сучасними методами збирання, обробки та аналізу статистичної інформації – невід’ємний елемент підготовки висококваліфікованих кадрів [1]. Основні прийоми та методи математичної статистики застосовують у сільському господарстві, промисловості, економіці тощо.

Ураховуючи, що основним завданням математичної статистики є не лише порівняння числових характеристик двох або більше груп, а й перспективний прогноз, що супроводжується аналізом та висновками в різних галузях, залишається актуальним представити можливе використання елементів математичної статистики у різних галузях.

Сутність проблеми обраної теми полягає у правильному створенні математичних моделей у сільському господарстві й промисловості та застосування методів математичної статистики до обраних галузей.

Метою роботи було описати основні методи математичної статистики та представити технологію застосування кореляційно-дисперсійного аналізу для характеристики утворених математичних моделей.

Завданням статистичної обробки експериментальних даних є представлення їх у формі, зручній для аналізу, та проведення такого аналізу з метою виявлення певних закономірностей у процесах, стан яких відображають експериментальні дані. У прикладному аспекті найчастіше статистична обробка даних спрямована на виявлення відмінностей між відповідними показниками та оцінку достовірності таких відмінностей [1]. Наприклад, на оцінку різниці між дослідною та контрольною групами об’єктів за певною ознакою та статистичною значимістю такої різниці [2]. Зрозуміло, що порівнюватись можуть не тільки дослідна й контрольна групи, а й, наприклад, групи об’єктів з різними властивостями.

Загальна кількість об’єктів, які мають деяку спільну ознаку, називається генеральною сукупністю. Деяка випадковим чином відібрана частина

генеральної сукупності називається вибірковою сукупністю або просто вибіркою. Об'єкти для дослідження, що довільним чином відбираються з генеральної сукупності, утворюють вибірку сукупність [3]. Кількість об'єктів генеральної сукупності називають обсягом генеральної сукупності. Число об'єктів вибіркової сукупності утворює обсяг вибірки. Якщо властивості об'єктів вибірки правильно відображають властивості об'єктів генеральної сукупності, то кажуть, що вибірка репрезентативна. Характеризуючи генеральну та вибірку сукупність, схиляються до обчислення таких числових характеристик, як середня арифметична величина, середнє квадратичне відхилення та дисперсія.

Середня арифметична – це узагальнююча величина, яка носить і абстрактний, і конкретний характер. Абстрактність середньої арифметичної полягає в тому, що вона може не дорівнювати за своїм числовим значенням жодному елементу вибірки. А конкретність виявляється в тому, що вона виражається в тих же одиницях виміру, що і варіанти вибірки. Її значення є середнім між найменшим та найбільшим значеннями вибірки [3].

Знаючи середнє арифметичне значення даних експерименту, виникає питання: як обчислити середню величину, на яку відрізняються дані від середнього арифметичного? Різницю між будь-яким виміром з вибірки й середнім арифметичним цієї ж вибірки називають відхиленням варіанти. Додаючи квадрати всіх різниць і ділячи на кількість цих різниць, отримуємо величину, яка називається дисперсією. Фактично вона показує середнє арифметичне квадратів відхилень. Для того, щоб позбутися квадрату величини, обчислюємо корінь квадратний з дисперсії. Отримане значення називається середнім квадратичним відхиленням [1].

Усі розглянуті показники варіації – середнє лінійне відхилення, середній квадрат відхилення та середнє квадратичне відхилення – завжди виражають у одиницях вихідних даних ряду та середньої величини [3]. Усі вони є абсолютним виміром варіації. А це означає, що безпосередньо порівнювати абсолютні показники варіації у варіаційних рядах різних явищ не можна. Для того щоб забезпечити їхнє порівняння, потрібно обчислити показники, які характеризують варіацію, виражену в стандартних величинах, наприклад у відсотках. Відношення абсолютних характеристик варіації до середньої величини називаються коефіцієнтами варіації.

Коефіцієнти варіації дають змогу порівнювати варіацію різних ознак або варіацію однієї ознаки у різних сукупностях. Для порівняння варіацій найчастіше використовують квадратичний коефіцієнт варіації. Цей показник вживається для оцінки однорідності сукупності, тобто надійності й типовості середньої величини.

Характеризуючи цілу сукупність лише за її частиною, неможливо уникнути помилок, які називаються помилками репрезентативності [2]. Навіть за ідеальної організації дослідницької роботи з'являються помилки такого типу. Помилка репрезентативності середньої арифметичної залежить від двох величин: від різноманітності ознаки в генеральній сукупності й від чисельності вибірки. Чим менша степінь різноманітності й чим більша кількість об'єктів

вибірки, тим меншою є величина помилки репрезентативності вибіркового середнього арифметичного [3].

Для розрахунку величини помилки використовується формула: $\Delta = \frac{\sigma}{\sqrt{n}}$,

де σ – середнє квадратичне відхилення, n – об'єм вибіркової сукупності [9].

Тепер будь-яку ознаку можна представити у формі: $x = \bar{X} \pm \Delta$, де \bar{X} – середнє арифметичне вибірки [2].

Якщо аналізу піддаються дві чи більше групи об'єктів, виникає питання оцінки різниці між групами за аналізованим показником. Найбільш поширеною є оцінка достовірності за методом Стюдента [1]. Для цього визначається

критерій достовірності різниці за формулою: $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\Delta_1^2 + \Delta_2^2}}$, де \bar{X}_1 і \bar{X}_2 –

середні значення в групах, Δ_1 і Δ_2 – помилки репрезентативності в групах. Обчислений критерій t порівнюється зі стандартним (табличним) значенням критерія Стюдента t_{st} для $n = n_1 + n_2 - 2$, де n_1 та n_2 – кількість вимірів у групах.

Для статистичної оцінки взаємозв'язків між явищами та їх істотності при невеликій кількості спостережень застосовують дисперсійний аналіз.

Дисперсійний аналіз – це метод оцінки впливу одного чи кількох факторів, що одночасно діють на певну результативну ознаку. Його застосовують під час статистичної обробки даних, одержаних унаслідок експерименту або спостереження, для виявлення впливу окремих факторів та їх взаємодії на рівень показників ефективності певної галузі.

Важливим поняттям є кореляція, яка виражає статистичну залежність між величинами і не має строго функціонального характеру [3]. Кореляційна залежність виникає тоді, коли одна з величин залежить не тільки від заданої другої, а й від деяких випадкових факторів; або коли серед умов, від яких залежать обидві величини, є загальні для них обох.

Кореляційний зв'язок може мати різну степінь – від повної незалежності до функціональної залежності. Крім того, характер зв'язку між різними величинами може бути різний. Тому виникає необхідність визначити форму, напрям і степінь кореляційних зв'язків та введення поняття коефіцієнта кореляції [3].

За формою кореляція може бути прямолінійною і криволінійною, за напрямком – прямою і оберненою. При додатній кореляції залежність між величинами буде прямою: при збільшенні однієї величини, збільшується й інша. При від'ємній кореляції залежність обернена: збільшення однієї величини пов'язано зі зменшенням другої. Ступінь кореляції вимірюється різними показниками зв'язку. Такими показниками є коефіцієнт кореляції, кореляційне відношення та ін.

Рівняння, за допомогою яких визначають статистичний зв'язок між корелюючими величинами, називають рівняннями регресії, а лінії, побудовані на їх основі, – лініями регресії [3].

Аналіз розглянутої методології підводить до практичного застосування кореляційно-регресивного методу, наприклад, у молочній промисловості. Вагомим внеском є представлені рівняння регресії для математичної моделі, що кількісно виражають зв'язок між фізико-хімічними характеристиками молока-сировини. Аналізуючи молоко трьох гатунків (екстра, вищий, перший), були встановлені кореляційні зв'язки між сезонністю і його фізико-хімічними показниками: масовою часткою жиру, білку та сухих речовин молока [4].

На основі знайдених коефіцієнтів кореляції були виведені рівняння лінійної регресії, які поєднують вміст білків, жирів та сухих речовин молока:

- для гатунку екстра: $y=0,356 x_1+3,571 x_2-0,015$;

- для вищого гатунку: $y=-0,21 x_1+4,397 x_2-0,01$;

- для першого гатунку: $y=1,229 x_1+2,654 x_2+0,03$,

де u, x_1, x_2 – масова частка відповідно сухих речовин, жиру, білка.

Рівняння регресії використовується для прогнозування очікуваних рівнів результативних ознак при встановлених значеннях факторних ознак. Таким чином, встановлені залежності, які поєднують основні фізико-хімічні показники молока-сировини залежно від його гатунковості. Це дасть змогу більш повно використовувати сировину у виробництві молочних продуктів.

Висновок: Для повної статистичної обробки експериментальних даних необхідно вибрати репрезентативну вибірку, для якої обчислити середнє значення та помилку репрезентативності. Для порівняння декількох груп і встановлення достовірності між ними користуються методом Ст'юдента. Зв'язок між досліджуваними ознаками представляють у вигляді коефіцієнта кореляції, а графічна інтерпретація тісноти демонструється рівнянням ліній регресії.

Отримані кореляційні зв'язки у молочній промисловості дають можливість більш повно використовувати молоко різних гатунків у технологічних процесів.

Література

1. Айвазян С.А. Прикладная статистика: Основы моделирования и первичной обработки данных / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1983. – 471 с.
2. Бешелев С.Л., Математико-статистические методы экспертных оценок / С.Л. Бешелев, Ф.Г. Гуревич. – М.: Статистика, 1980. –159 с.
3. Гаркавий В.Г. Математична статистика / В.Г. Гаркавий, В.В. Ярова. – К: Професіонал, 2004. – 484 с.
4. Технологія незбираномолочних продуктів: Навчальний посібник / Т.А. Скорченко, Г.Є. Поліщук, О.В. Грек, О.В. Кочубей – Вінниця : Нова книга, 2005. – 264 с.