

ПРО ОДНОФАКТОРНИЙ ТА ДВОФАКТОРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ

Всі процеси, що відбуваються у природі чи людському суспільстві, є наслідком взаємодії багатьох факторів. Для того щоб вивчити ці процеси і надалі керувати ними, необхідно з'ясувати, яку роль у досліджуваному процесі відіграє кожний фактор окремо. Наприклад, у разі вивчення руху тіла слід з'ясувати, які сили спричинюють його рух, а які гальмують; яким чином саме рухоме тіло впливає на ті сили, що діють на нього. Досліджуючи процес зміни курсу деякої валюти, скажімо гривні, потрібно з'ясувати вплив багатьох економічних і соціальних факторів як внутрішніх, так і зовнішніх, що можуть істотно змінювати курс національної валюти щодо долара, німецької марки і т. ін. Усі зазначені фактори необхідно подати з допомогою певних кількісних оцінок, а далі – скористатися відповідними математичними методами. Отже, щоб мати змогу застосувати математичні методи з метою вивчення взаємодії тих чи інших факторів, слід уміти виражати дію кожного з них кількісно.

Щоб дістати потрібні числові дані, необхідно провести серію спостережень. Отже, спостереження є найважливішою ланкою будь-якого експерименту. Слід, проте, враховувати, що жодний найретельніше підготовлений експеримент не дозволяє виокремити саме той фактор, який для нас головний. Адже в здійснюваному експерименті ми не в змозі вилучити численні зайві фактори, які нас не цікавлять. Так, вивчаючи падіння тіла, ми не уникнемо дії на нього сил, зумовлених обертанням Земної кулі. Коли ж ідеться про хімічні реакції, нам ніколи не доведеться стикатися з чистими елементами. А досліджуючи вплив на врожайність тієї чи іншої культури внесеного в ґрунт добрива, ми не можемо знехтувати впливом інших факторів (опадів, середня весняна температура, економічний стан регіону і т. ін.), які безпосередньо впливають на остаточний наслідок експерименту – урожайність.

Отже, кожне спостереження дає нам лише наслідок взаємодії основного фактора, який нас цікавить, з багатьма сторонніми, другорядними. Деякі з них погрібне й можна враховувати в дослідженнях. Врахування ж решти факторів або в принципі неможливе, або недоцільне з якихось міркувань. Тому за реальних умов під час дослідження будь-якого процесу застосовують метод його формалізації, беручи до уваги-інше ті фактори, які істотно впливають на зазначений процес.

Так настають непередбачені наперед події, котрі називають випадковими. Випадкові події в масі спостережень підпорядковані, як з'ясували дослідники, певним характерним лише для них не випадковим законам. Математична наука, що вивчає закономірності масових подій називається теорією ймовірностей.

Науку, що використовує теорію ймовірностей для обробки численних одиниць інформації як наслідків експерименту, називають математичною статистикою.

1. Загальні відомості про дисперсійний аналіз

Дисперсійний аналіз був створений спочатку для статистичної обробки агрономічних дослідів. В наш час його також використовують як в економічних експериментах, так і технічних, соціальних.

Сутність цього аналізу полягає в тому, що загальну дисперсію досліджуваної ознаки розділяють на окремі компоненти, які обумовлені впливом певних конкретних чинників. Істотність їх впливу на цю ознаку здійснюється методом дисперсійного аналізу. Відповідно до дисперсійного аналізу будь-який його результат можна подати у вигляді суми певної кількості компонент.

Складнішою моделлю аналізу є вивчення впливу на результати експерименту кількох факторів. У разі проведення дисперсійного аналізу досліджуваний масив даних, одержаних під час експерименту, поділяють на певні групи, які різняться дією на результати експерименту певних рівнів факторів. Вважається, що досліджувана ознака має нормальний закон розподілу, а дисперсії в кожній окремій групі здобутих значень ознаки однакові. Ці припущення необхідно перевірити.

2. Однофакторний дисперсійний аналіз

Нехай потрібно дослідити вплив на ознаку X певного одного фактора. Результати експерименту ділять на певне число груп, які відрізняються між собою ступенем дії фактора. Для зручності в проведенні необхідних обчислень результати експерименту зводять в спеціальну таблицю.

Відповідно до моделі однофакторного дисперсійного аналізу необхідно визначити дні дисперсії, а саме: міжгрупову (дисперсію групових середніх), зумовлену впливом досліджуваного фактора на ознаку X , і внутрішню групову, зумовлену впливом інших випадкових факторів. Загальна дисперсія розглядається як сума квадратів відхилень.

Для того щоб мати виправлені дисперсії, необхідно кожену зі здобутих сум поділити на число ступенів свободи. Виправлена дисперсія, що характеризує розсіювання всередині групи, зумовлене впливом випадкових факторів.

Виправлена дисперсія, що характеризує розсіювання групових середніх відносно загальної середньої, яке викликане впливом фактора на результат експерименту ознаки X .

Завдання виявлення впливу фактора на наслідки експерименту полягає в порівнянні виправлених дисперсій. І справді, якщо досліджуваний фактор не впливає на значення ознаки X , то в цьому разі виправлені дисперсії можна розглядати як незалежні оцінки загальної дисперсії D . І навпаки, якщо відношення їх істотне, то в цьому разі вибірки слід вважати здійсненими з різних сукупностей, тобто з сукупностей з різним рівнем впливу фактора.

Порівняння двох дисперсій ґрунтується на перевірці правильності нульової гіпотези: – про рівність дисперсій двох вибірок.

За статистичний критерій вибирається випадкова величина, що має розподіл Фішера-Снедекора з $k_1 = N - p$ та $k_2 = p - 1$ ступенями свободи. За цими значеннями знаходимо критичну точку $F_{кр}$.

Якщо $F^* < F_{кр}$, то нульова гіпотеза про вплив фактора на результати досліджень відхиляється, а коли $F^* > F_{кр}$, то цим самим підтверджується вплив фактора на ознаку X .

3. Двофакторний дисперсійний аналіз

Нехай необхідно визначити вплив двох факторів A і B на певну ознаку X . Для цього необхідно, щоб дослід здійснювався при фіксованих рівнях факторів A і B , а також їх одночасній дії на ознаку. При цьому дослід здійснюватимемо n разів для кожного з рівнів факторів A і B .

Позначимо через x_{ijk} конкретне значення ознаки X , якого вона набуває при i -му експерименті, j -му рівні фактора A і k -му рівні фактора B .

1) $F_A^* > F_{кр}$, то нульова гіпотеза про відсутність впливу фактора A відхиляється;

2) $F_B^* > F_{кр}$, то нульова гіпотеза про відсутність впливу фактора B відхиляється;

3) $F_{AB}^* > F_{кр}$, то нульова гіпотеза про відсутність спільного впливу факторів A та B відхиляється.

Отже, дисперсійний аналіз створений для статистичної обробки агрономічних дослідів, а також його використовують в економічних експериментах, технічних, соціальних. Сутність цього аналізу полягає в тому, що загальну дисперсію досліджуваної ознаки розділяють на окремі компоненти, які обумовлені впливом певних конкретних чинників.

Література:

1. Жлуктенко В.І., Наконечний С.І. Теорія ймовірностей і математична статистика: Навч. – метод. посібник. У 2 ч. – Ч I. Теорія ймовірностей. – К.: КНЕУ, 2000. – 347 с.

2. Жлуктенко В.І., Наконечний С.І., Савіна С.С. Теорія ймовірностей і математична статистика: Навч. – метод. посібник. У 2 ч. – Ч II. Математична статистика. – К.: КНЕУ, 2001. – 336 с.

3. Турчин В.М. Теорія ймовірностей: Основні поняття, приклади, задачі: Навч. посіб. – К.: Видавництво А.С.К., 2004. – 208 с.

4. О.В. Крайчук, Г.К. Московська, О.І. Соколенко. Теорія ймовірностей і математична статистика. – Рівне, 2004.

5. Зубков А.М., Севастьянов Б.А., Чистяков В.В. Сборник задач по теории вероятностей. – М.: Наука, 1989. – 320 с.