УДК 519.246

Величко В.Є., Батуніна В.П.

[1] *канд. фіз.-мат. наук, доцент кафедри алгебри, СДПУ*

[2] *асистент кафедри алгебри, СДПУ*

e-mail: vladislav.velichko@gmail.com

СТАТИСТИЧНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТУ В ЕЛЕКТРОННИХ ТАБЛИЦЯХ OpenOffice.org Calc

Статистичний аналіз результатів будь-якого експерименту передбачає виконання певної кількості обчислень. Зрозуміло, що такі обчислення зручно виконувати в електронних таблицях. Крім того вони містять велику кількість стандартних статистичних функцій і їх використання буде корисним для статистичного аналізу результатів експерименту.

Ключові слова: *математична статистика, результати експерименту, електронні таблиці*

Вступ

Аналіз результатів експерименту за допомогою математичної статистики зазвичай зводиться до перевірки справедливості припущень, або гіпотез, відносно вивчаємого фізичного явища яке задається отриманими під час експерименту даних. Наприклад, до перевірки припущень про рівносильність результатів виміру однієї і тієї ж постійної величини, якщо виміри виконуються двома незалежними дослідниками на різних установках.

Гіпотезою, яку належить перевірити, може стати правомірність використання фізичної моделі, яку було вибрано для описання експерименту. Оскільки модель дозволяє теоретично передбачити вид функціонального зв'язку між величинами, які вимірюються, то статистичний аналіз експериментальної залежності, який проводиться з врахуванням висновків моделі, дає інформацію про те, чи достатньо справедливий модельний опис.

Основна частина

Покладемо, що даний статистичний розподіл вирівняний за допомогою деякої теоретичної кривої $f(x)$. Як би не гарно не була підібрана теоретична крива, між нею і статистичним розподілом неминучі деякі розбіжності.

Природно постає питання: пояснюються ці питання тільки випадковими обставинами, які пов'язані з обмеженістю кількості спостережень, або вони є суттєвими і пов'язаними з тим, що підібрана крива погано вирівнює цей статистичний розподіл. Відповіддю на поставлене питання є так званий критерій узгодженості.

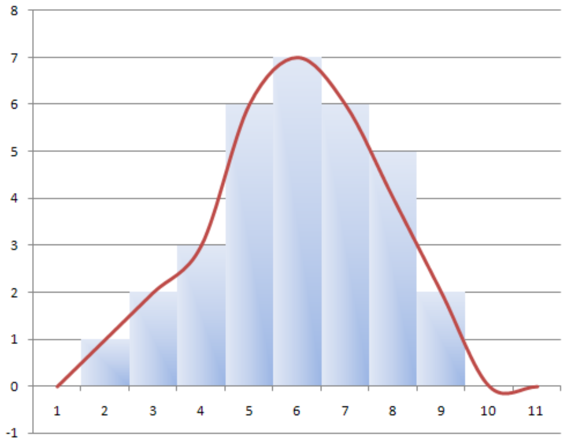

Рис. 1: Статистичний розподіл та теоретична крива $f(x)$

На основі даного статистичного матеріалу перевіряють гіпотезу H, яка полягає в тому, що випадкова величина X підпорядковується деякому закону розподілу $F(x)$. Для того, щоб прийняти або спростувати гіпотезу H, розглянемо деяку величину U, яка характеризує степінь розходження теоретичного і статистичного розподілів. Величина U може бути задана різноманітними способами. Тим не менш вона є випадковою величиною і закон розподілу цієї випадкової величини залежить від закону розподілу випадкової величини X, над якою виконувались дослідження, і від числа досліджень n. Якщо гіпотеза H вірна, то закон розподілу величини U визначається законом розподілу величини X і числом n.

Розглянемо один з багатьох критеріїв узгодженості який найчастіше використовується – «критерій χ^2» Пірсона. Припустимо, що виконано n незалежних дослідів, в кожному з яких випадкова величина X прийняла деяке значення. Результати дослідів зведені в k розрядів і оформлені у вигляді статистичного ряду:

I_i	$x_1;x_2$	$x_2;x_3$	$\dots$	$x_k;x_{k+1}$
p_i^*	p_1^*	p_2^*	$\dots$	p_k^*

Табл. 1: Розподіл випадкової величини за інтервалами.

Необхідно перевірити, чи узгоджуються експериментальні дані з гіпотезою про те, що випадкова величина X має даний закон розподілу заданий функцією розподілу $F(x)$. Назвемо цей закон розподілу теоретичним. Знаючи теоретичний закон розподілу, можна знайти теоретичні імовірності попадання випадкової величини в кожен з розрядів: $p_1, p_2, \dots, p_k$. Перевіряючи узгодженість теоретичного і статистичного розподілів, будемо виходити з розбіжностей між теоретичними ймовірностями p_i і частотами які отримали в результаті експерименту p_i^*. К. Пірсон показав, що при великих значеннях n закон розподілу

$$U = \chi^2 = n \sum_{i=1}^{k} \frac{(p_i^* - p_i)^2}{p_i}$$

практично не залежить від функції розподілу $F(x)$ і від числа дослідів n, а залежить тільки від числа розрядів k, зокрема, цей закон при збільшенні n наближається до так званого розподілу χ^2 [1].

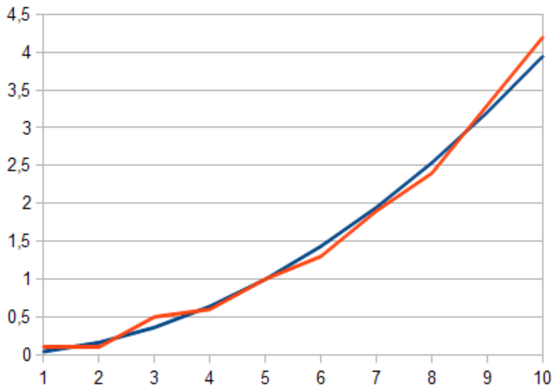

Рис. 2: Статистична крива та теоретична крива $f(x)$

Розподіл χ^2 залежить від параметру r, який називають числом «степенем свободи» і яке рівне кількості розрядів k мінус число незалежних умов («зв'язків»), які накладені на частоти p_i^*. Прикладами таких умов можуть бути: сума ймовірностей рівна одиниці (таку умову накладають завжди), рівність математичних сподівань, рівність дисперсій тощо. Для розподілу χ^2 складені спеціальні таблиці, користуючись якими можна для кожного значення χ^2 і числа степенів свободи r знайти імовірність p того, що величина, яка розподілена за цим законом, перебільшить це значення.

Таким чином, схема використання критерію χ^2 до оцінки узгодженості теоретичного і статистичного розподілу наступна:

1. Визначається міра розходження χ^2 за наведеною формулою.

2. Визначається число степенів свободи r як $r = k - s$.

3. За r і χ^2 за допомогою таблиці визначається імовірність того, що величина, яка має розподіл χ^2 з r степенями свободи, перевершить дане значення χ^2. Якщо ця імовірність мала, гіпотеза відбраковується як неверна. Якщо імовірність велика, гіпотезу можна вважати такою, що не протирічить дослідним даним.

Слід зазначити, що питання про те, якою повинна бути імовірність, щоб відхилити гіпотезу не визначений і не може бути розв'язаний за допомогою математичних міркувань.

Одним із засобів виконань обчислень з великою кількістю даних є електронні таблиці. Це один із зручних засобів обчислень та аналізу даних, в тому числі й статистичного аналізу. Наявність графічного представлення даних в електронних таблицях тільки покращує отримані результати.

Електронні таблиці OpenOffice.org Calc відносяться до програмного забезпечення вільного користування, тим не менш це програмне забезпечення містить 57 вбудованих статистичних функцій, за допомогою яких можна розв'язувати досить складні статистичні задачі.

Однією з таких задач і є задача визначення критерію узгодженості статистичних експериментальних даних та теоретичного розподілу. Візьмемо, для прикладу, результати деякого експерименту. Нехай ці результати залежать від параметра, який впливає на отриманий результат. Виконавши виміри значень по декілька раз для кожного параметру і обчисливши середнє арифметичне розбиваємо на інтервали за параметром і таким чином отримуємо розподіл випадкової величини. Якщо, для прикладу побудувати графік теоретичних значень та статистичних значень то можна отримати «майже схожі» графіки (рис 2).

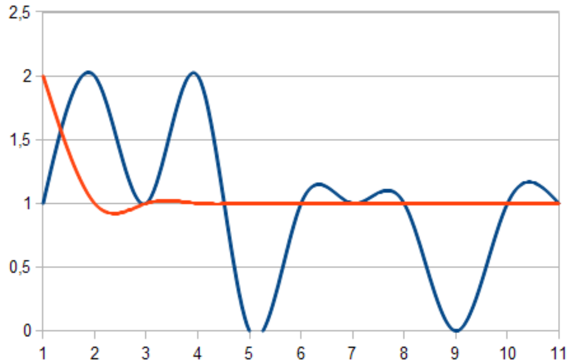

Рис. 3: Статистична крива розподілу та теоретична крива розподілу $f(x)$

Для обчислення критерію χ^2 в електронних таблицях знайдемо максимальне (функція **MAX**) та мінімальне (функція **MIN**) значення нашого інтервалу. Поділимо інтервал значень на певну кількість підінтервалів і за допомогою функції **COUNTIF** з параметрами, що будуть задовольняти побудованим інтервалам, підрахуємо кількість значень які входять до кожного з підінтервалів. Отриману так звану функцію розподілу випадкової величини або ще кажуть – функцію частот випадкової величини. Поділивши отримані значення на загальну кількість вхідних даних отримуємо так звану функцію щільності статистичного розподілу. Аналогічні дії виконуємо і для нашої теоретичної кривої, тобто функції яка максимально близько описує отримані результати, на тих самих значеннях аргументу або параметру на яких отримували статистичні значення. Отримуємо дві функції розподілу випадкової величини (рис 3.) За наведеною формулою критерію Пірсона обчислюємо величину χ^2 та число степенів свободи. Потім за допомогою функції **CHIDIST** за обчисленим значенням χ^2 та числом спетенів свободи обчислюємо імовірність відповідності статистичних даних та теоретичної кривої. Отримана імовірність і буде свідчити про те, наскільки наше припущенна про рівносильність статистичних даних та значень теоретичної кривої.

Слід зазначити, що у випадку, коли буде достатнім наявність тільки одного зв'язку, тобто $r = k-1$, можна скористатись стандартною функцією **CHITEST**, чого інколи буде і достатньо. В разі коли ми накладаємо додаткові умови (тобто «зв'язки») необхідно скористатись вищенаведеним способом обчислення за допомогою електронних таблиць.

Висновки

Розвиток інформаційних технологій, а особливо технологій обробки числової інформації дозволяє проводити статистичний аналіз отриманих експериментальних даних, що в свою чергу покращує не тільки наочні результати експерименту але і аналізує якість його проведення. Наявність в електронних таблицях OpenOffice.org Calc досить значної кількості вбудованих статистичних функцій дозволяє виконувати нескладний статистичний аналіз, який буде корисним для осмислення отриманих даних будь-якого експерименту.

Література

[1] *Karl Pearson* On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. – Philosophical Magazine, Series 5 50 (302). – P. 157 – 175.

[2] *Кендалл М.* Статистические выводы и связи / М. Кендалл, А. Стьюарт. – М.: Наука, 1973. – 900 с.

[3] База Знаний: Calc. Статистические функции // Инфра-Ресурс. Дата обновления: 28.01.2012
URL: http://wiki.i-rs.ru/wiki/RU/kb/module/calc/function/statistical