

## **Перевірка статистичних гіпотез. Критерій хі-квадрат**

**Критерій  $\chi^2$**  Нехай по деяким даним (спостереженням) було підбрано розподіл, що має  $m$  параметрів (наприклад, гауссівський розподіл має 2 параметри) і необхідно перевірити придатність підбору (тобто, чи не суперечать ці дані пропонованому-підбраному розподілу). Для цього область можливих значень спостережень розбивають на  $k$  несумісних інтервалів (клітин). Кожне із спостережень має потрапити до однієї з клітин. Нехай маємо  $n$  спостережень. Очікувана кількість спостережень для кожної клітини обраховується на основі обраного розподілу. В результаті складаємо таку таблицю:

|                                      | Кількості спостережень у відповідних клітинах: |       |     |       | Всього |
|--------------------------------------|------------------------------------------------|-------|-----|-------|--------|
|                                      | 1                                              | 2     | ... | k     |        |
| Очікувана кількість<br>(expected)    | $e_1$                                          | $e_2$ | ... | $e_k$ | $n$    |
| Спостережена<br>кількість (observed) | $o_1$                                          | $o_2$ | ... | $o_k$ | $n$    |

При побудові цієї таблиці слід обирати інтервали так, щоб в кожен з них потрапило хоч одне спостереження (в деяких книжках радиться, щоб до кожного інтервалу потрапило не менше 4 спостережень). Відзначимо, що розподіл ймовірностей по клітинах має поліноміальний розподіл незалежно від розподілу, що перевіряється.

Гіпотеза  $H_0$ :  $p_i = e_i/n$  ( $i = 1, \dots, k$ ), де  $p_i$  – ймовірність кожного спостереження потрапити до клітини №  $i$ . Альтернативна гіпотеза  $H_1$ :  $\bar{H}_0$ . Критеріальна статистика:

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}.$$

Нехай  $\chi^2_{1-\alpha}(k-m-1)$  квантиль розподілу  $\chi^2$  із  $(k-m-1)$  ступенями вільності та індексом (довірчою ймовірністю)  $p=1-\alpha$  (цей квантиль шукають за таблицями). Якщо  $\chi^2 < \chi^2_{1-\alpha}(k-m-1)$ , то гіпотезу  $H_0$  приймаємо, інакше – відхиляємо. Якщо параметри розподілу пропонуються заздалегідь, а не оцінюються за вибіркою, то беруть  $\chi^2_{1-\alpha}(k-1)$  квантиль.

**Задача 11.1** Двадцять спостережень взяті із невідомої сукупності:

0,33; -0.52; -2.41; -1.93; 0.46; -0.44; -0.97; -0.38; 0.48; 1.29;  
-1.82; -1.23; -0.21; 2.66; -1.22; -0.41; -0.95; 1.47; -0.83; -0.43.

Перевірити гіпотезу  $H_0$  про те, що розглянутий розподіл є стандартним гауссівським (з рівнем значимості 5%).

**Задача 11.2** Розв'язати задачу 11.1 при умові, що параметри розподілу були отримані в результаті оцінювання за вибіркою.

**Задача 11.3** Було перевірено 11 ручних гранат. Потрібно було перевірити твердження виробника про те, що середній час спрацьовування вибухівки складає 4,01 с. При перевірці отримані такі дані:

4,21; 4.03; 3.99; 4.05; 3.89; 3.98; 4.01; 3.92; 4.23; 3.85; 4.20.

Зробити статистичний висновок для рівня значимості 10%.

**Задача 11.4** Автомобіль на 100 км витрачає 10л пального. Із новим двигуном очікується зменшення споживання пального. Перевірено 25 нових автомобілів із новим двигуном і отримали  $\bar{x}=9,3$  л.

а) Вважаючи розподіл нормальним із  $\sigma^2 = 4$  л<sup>2</sup> перевірити гіпотезу про те, що новий двигун не економить паливо, тобто має такі ж витрати, як і старий (рівень значимості  $\alpha = 5\%$ ).

б) Знайти ймовірність похибки першого роду.

**Задача 11.5** Нехай в умовах попередньої задачі разом з гіпотезою  $H_0: m=10$  розглядається альтернативна гіпотеза  $H_1: m=9$ . Побудувати критичну область, обравши в якості критерію вибіркове середнє. Знайти ймовірності похибок першого та другого роду.

**Задача 11.6** Який мінімальний об'єм  $n$  вибірки слід взяти в попередній задачі, щоб ймовірність похибки першого роду склала 1%, а ймовірність похибки другого роду не перевищувала 10%. Якою буде критична область для вибіркового середнього?

### **Домашнє завдання № 11**

1. Розв'язати задачу 11.5 для критичної області  $\bar{X} < 9.5$ .
2. Монету підкинули 100 раз і при цьому 67 раз випав герб. Перевірити гіпотезу про те, що монета симетрична (при рівні значимості 5%).