

$\chi^2$  - тестове.

- тест за принадлежност към дадено разпределение: Ако  $(X_1, \dots, X_k)$  е многомерно разпределена случайна величина с параметри  $n, p_1, \dots, p_n$ ,

то за голямо  $n$  случайната величина  $\sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i}$  има приблизително  $\chi^2$  разпределение с  $k - 1$  степени на свобода.

Пример: Наблюдавани са 150 случая на срив на компютърна система и всеки от тях е класифициран в една от четири категории: софтуерен проблем - 13, хардуерен проблем - 10, потребителска грешка - 42 и претоварване на системата - 65. Предварително се смята, че тези причини имат следните честоти: 10%, 5%, 25%, 40%. Потвърждават ли данните това предположение? (Наблюдаваната стойност на  $\chi^2$  статистиката е 5.39, откъдето p-value е 0.25.)

- тест за еднородност за  $2 \times 2$  честотна таблица, където  $n_{i\bullet}$  са фиксирани предварително:  $\sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n})^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}}$  има приблизително  $\chi^2$  разпределение с 1 степен на свобода.

Пример: изследва се заболяемостта от определен вид болест при изложените на радиоактивност, като са изследвани 300 души, изложени на радиоактивно влияние и 320 - неизложени на такова влияние. Тук се сравняват пропорциите на заболяелите в двете групи. (ако имаме 52 болни от изложените на радиация и 48 от неизложените, стойността на тестовата статистика е 0.62, което не е значимо дори за  $\alpha = 0.25$ )

- тест за независимост за  $2 \times 2$  честотна таблица:  $\sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n})^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}}$  има приблизително  $\chi^2$  разпределение с 1 степен на свобода.

Пример: В проучване на зависимостта на заболяването от рак на белия дроб и излагането на влиянието на азбестов прах са получени следните резултати: 50 от 5000 души са развили рак, 10 от тях са имали досег с азбест, като общо 500 от изследваните са имали досег. (Наблюдаваната стойност на  $\chi^2$  статистиката е 5.61, откъдето p-value е между 0.01 и 0.025.)

За  $r \times c$  честотна таблица, съответната статистика има  $\chi_{(r-1)(c-1)}^2$  разпределение.