Статистична перевірка гіпотез.

Приймаючи рішення в умовах невизначеності, ми змушені висловлювати деякі гіпотези або припущення і доводити їх з певним ступенем надійності.

Статистичною називають гіпотезу про властивості або ознаки генеральної сукупності, яка перевіряється на основі вибірки.

У математичній статистиці виділяють два основних види статистичних гіпотез:

1) гіпотези про закон розподілу ймовірностей випадкової величини (ознаки генеральної сукупності);

2) гіпотези про значення параметрів розподілу випадкової величини (ознака генеральної сукупності).

Статистичні гіпотези першого типу називаються непараметричних, а другого типу - параметричними.

Нульовою (основною) називають висунуту гіпотезу H_0.

Конкуруючої або альтернативної називають гіпотезу H_1, яка суперечить нульовий.

$H_0 : a = 10$

$H_1 : a \neq 10$

Простою називають гіпотезу, яка містить лише одне припущення. **Складною** називають гіпотезу, яка складається з кінцевого або нескінченного числа простих гіпотез.

У результаті статистичної перевірки гіпотези може бути прийнято одне з двох правильних рішень:

1) гіпотеза приймається і вона істинна;

2) гіпотеза відхиляється і вона не істинна.

Поряд з цим в результаті статистичної перевірки гіпотез, можуть бути допущені помилки (прийняті неправильні рішення) 2-х типів:

1) гіпотеза відхилена, але вона істинна (помилка першого роду);

2) гіпотеза приймається, але вона не істинна (помилка другого роду).

Імовірність припуститися помилки першого роду називається **рівнем значущості** і позначається: $P(H_1/H_0) = \alpha$. Число α задається наперед і звичайно дорівнює 0.1; 0.05; 0.01.

Статистичним критерієм називають випадкову величину K, яка служить для перевірки нульової гіпотези. K обирають такою, щоб її закон розподілу був відомий.

$K_{наблюд.}$ - значення, що обчислено за вибіркою.

Критичною областю називають сукупність значень критерію K, за яких нульову гіпотезу H_0 відхиляють.

Областю прийняття гіпотези або областю допустимих значень називають сукупність значень критерію K, за яких гіпотезу H_0 приймають.

Правило перевірки гіпотез:

• якщо спостерігається значення критерію $K_{наблюд.}$ належить критичної області – гіпотезу H_0 відхиляють;

• якщо спостерігається значення критерію $K_{наблюд.}$ належить області прийняття гіпотези – гіпотезу H_0 приймають.

Критичними точками (межами) $K_{кр.}$ називають точки, що відокремлюють критичну область від області прийняття гіпотези

Правостороння критична область: $K > K_{кр.}$, $K_{кр.} > 0$.

Лівостороння критична область: $K < K_{кр.}$, $K_{кр.}$<0.

Односторонньою називають правосторонню або лівосторонню критичну область.

Двусторонньою називають критичну область: $K < K_1; K > K_2; K_2 > K_1$.

Потужністю критерію називають імовірність належності критерію критичній області за умови, що істина конкуруюча гіпотеза.

Схема перевірки статистичної гіпотези:

1. формулювання гіпотези H_0, гіпотези H_1, завдання рівня значущості α для перевірки H_0;

2. визначення критерію K для перевірки гіпотези H_0, який є випадковою величиною з відомим розподілом ймовірності;

3. визначення критичної області щодо даного критерію K і рівня значущості α, знаходження $K_{кр.}$;

4. знаходження $K_{наблюд.}$ на підставі даних конкретної вибірки;

5. прийняття рішення: якщо $K_{наблюд.}$ потрапляє в критичну область - гіпотезу H_0 відхиляють; якщо $K_{наблюд.}$ потрапляє в область допустимих значень, гіпотезу H_0 приймають.

Перевірка непараметричних гіпотез.

Перевірка гіпотези про передбачуване законі невідомого розподілу проводиться за допомогою спеціально підібраною випадкової величини - критерію згоди.

Критерієм згоди називається критерій перевірки гіпотези про передбачуване законі невідомого розподілу.

Існує кілька критеріїв згоди:

• критерій Пірсона;

• критерій Колмогорова;

• критерій Смирнова.

Критерій χ^2 Пірсона дозволяє перевіряти гіпотезу про узгодження даних вибірки з конкретними теоретичними розподілами для будь випадкової величини (тобто з будь-яким законом розподілу, безперервної і дискретної).

Критерій χ^2 Пірсона:

Введемо статистику:

$$K = \chi^2 = \sum_{i=1}^{m} \frac{\left(n_i - n_i'\right)^2}{n_i'} \quad (1),$$

де m - число груп у статистичному розподілі, n_i – емпірична частота ознаки X в i- ій групі, n_i' – теоретична частота.

Якщо нульова гіпотеза істина, то при $n \to \infty$ закон розподілу статистики χ^2 прямує до χ^2 розподілу з числом ступенів свободи $k = m - s - 1$, де m - число інтервалів, а s - число параметрів гіпотетичної функції $F(x, \theta_j)$, що оцінюються за основі даних спостережень.

Обчислив за формулою (1) $K_{эмп.} = \chi^2{}_{эмп.}$ і визначивши $\chi^2_{кр.}(\alpha, k) = K_{кр.}$ (використовуючи таблицю) роблять такі висновки:

1. якщо $K_{эмп.} \geq K_{кр.}$ гіпотезу H_0 відхиляють;
2. якщо $K_{эмп.} < K_{кр.}$ гіпотезу H_0 приймають.