Непараметричні тести значущості

Формули

Непараметричні тести значущості - поточкова схема

1. Сформулюємо основну гіпотезу H_0, що стосується:
   - відповідність генеральної сукупності деякому розподілу, або:
   - вибіркова випадковість
2. Розраховуємо відповідну статистику.
3. Створюємо і рисуємо критичну область (дво або односторонню залежно від H_1)
4. Перевіряємо, чи знаходиться статистика в критичній області. Якщо так - ми відхиляємо гіпотезу H_0 на користь альтернативної гіпотези H_1. Якщо ні, ми робимо висновок, що немає підстав для відхилення гіпотези H_0.

I. Тест на сумісність Пірсона

1. Формулюємо гіпотези:

H_0 : генеральна сукупність має розподіл …

H_1 : загальна сукупність не має такого розподілу

2. Визначаємо статистику: $\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}$

де r - кількість інтервалів у ряді, n_i емпіричні числа у вибірці, p_i теоретичні ймовірності/відсотки, n загальна кількість вибірки, np_i теоретичні числа

- Ймовірності p_i в нормальному розподілі визначаємо, правильно читаючи таблиці
- Ймовірності p_i в розподілі Пуассона визначаємо за формулою: $p_i = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$, де λ – середнє значення розподілу (тут зазвичай беремо середнє значення вибірки $\overline{X}$)
- Ймовірності p_i розподіді Бернуллі/до бінома обчислюємо за формулою: $p_i = \binom{n}{x_i} p^{1-x_i} (1-p)^{x_i}$, де p це ймовірність успіху з однієї вибірки.

3. Створюємо і малюємо праву критичну область для розподілу хі-квадрат, для $r-k-1$ ступенів свободи, де k позначає кількість параметрів у теоретичному розподілі ($k=2$ у нормальному розподілі, $k=1$ в розподілах Пуассона та біноміальних/Бернуллі)

4. Перевіряємо, чи статистика в критичній області. Якщо так - ми відхиляємо гіпотезу H_0 на користь альтернативної гіпотези H_1. Якщо ні, ми робимо висновок, що немає підстав для відхилення гіпотези H_0.

I. Тест на випадковість вибірки

II.a Для невеликого розміру вибірки

1. Формулюємо гіпотезу:

H_0: вибірка має випадковий характер

H_1: загальна сукупність не має такого розподілу

2. Ми впорядковуємо вибірку в порядку зростання і призначаємо літеру *a* всім результатам, якщо вона менша за медіану; *b*, якщо більше (якщо дорівнює – пропустити)

3. Ми повертаємо результати до рядка *aababb*… . Позначимо кількість рядів через *k*. Позначимо номери символів *a* і *b* через n_1 та n_2.

4. Зчитуємо граничні значення з таблиць розподілу рядів k_1 і k_2 такі, щоб $P\left(k \le k_1\right) = \frac{\alpha}{2}$; $P\left(k \le k_2\right) = 1 - \frac{\alpha}{2}$.

5. Перевіряємо, чи належить *k* критичній області $k < k_1 \vee k > k_2$ та записуємо відповідь.

II.b Для великого розміру вибірки

1. Формулюємо гіпотези:

H_0 : генеральна сукупність має розподіл …

H_1 : загальна сукупність не має такого розподілу

2. Ми впорядковуємо вибірку в порядку зростання і призначаємо літеру *a* всім результатам, якщо вона менша за медіану; *b*, якщо більше (якщо дорівнює – пропустити).

3. Ми повертаємо результати до рядка *aababb*… . Позначимо кількість рядів через *k*. Позначимо номери символів *a* і *b* через n_1 і n_2.

4. Визначаємо статистику: $Z = \dfrac{k - \left(\dfrac{2n_1 n_2}{n_1 + n_2} + 1 \right)}{\sqrt{\dfrac{2n_1 n_2 \left(2n_1 n_2 - n_1 - n_2\right)}{\left(n_1 + n_2\right)^2 \left(n_1 + n_2 - 1\right)}}}$,

5. Створюємо двосторонню критичну область для нормального розподілу.

6. Перевіряємо, чи статистика в критичній області. Якщо так - ми відхиляємо гіпотезу H_0 на користь альтернативної гіпотези H_1. Якщо ні, ми робимо висновок, що немає підстав для відхилення гіпотези H_0.

II. Тест на сумісність двох дистрибутивів

III.a Для невеликого розміру вибірки

1. Формулюємо гіпотези:

H_0 : вибірки надходять з однієї сукупності (з однаковим розподілом)

H_1 : вибірки з однієї сукупності

2. Результати обох вибірок упорядковуємо в послідовність без спаду (у разі однакових значень в обох вибірках спочатку записуємо результати з першого, потім з другого) і всім приписуємо букву *a* результати першого зразка та всі результати другого зразка до букви *b*.

3. Позначимо кількість рядів через *k*. Позначимо номери символів *a* і *b* через n_1 і n_2 .

4. З таблиць послідовного розподілу читаємо граничне значення k_1 так, щоб $P\left(k \le k_1\right) = \alpha$ (ліва критична область).

5. Перевіряємо, чи належить *k* критичній області $k \le k_1$ та записуємо відповідь.

III.b Для великого розміру вибірки

1. Формулюємо гіпотези:

H_0 : вибірки надходять з однієї сукупності (з однаковим розподілом)

H_1 : вибірки з однієї сукупності

2. Результати обох вибірок упорядковуємо в послідовність без спаду (у разі однакових значень в обох вибірках спочатку записуємо результати з першого, потім з другого) і всім приписуємо букву *a* результати першого зразка та всі результати другого зразка до букви *b*.

3. Визначаємо статистику: $Z = \dfrac{k - \left(\dfrac{2n_1 n_2}{n_1 + n_2} + 1 \right)}{\sqrt{\dfrac{2n_1 n_2 \left(2n_1 n_2 - n_1 - n_2\right)}{\left(n_1 + n_2\right)^2 \left(n_1 + n_2 - 1\right)}}}$,

4. Створюємо двосторонню критичну область для нормального розподілу.

5. Перевіряємо, чи статистика в критичній області. Якщо так - ми відхиляємо гіпотезу H_0 на користь альтернативної гіпотези H_1. Якщо ні, ми робимо висновок, що немає підстав для відхилення гіпотези H_0.