

Nieparametryczne Testy Istotności

Wzory

Nieparametryczne testy istotności – schemat postępowania punkt po punkcie

1. Formułujemy hipotezę główną H_0 odnoszącą się do:
 - zgodności populacji generalnej z jakimś rozkładem, lub:
 - losowości próby
2. Obliczamy odpowiednią statystykę.
3. Tworzymy i rysujemy obszar krytyczny (obu lub jednostronny w zależności od H_1)
4. Sprawdzamy, czy statystyka znalazła się w obszarze krytycznym. Jeśli tak – odrzucamy hipotezę H_0 na rzecz hipotezy alternatywnej H_1 . Jeśli nie – stwierdzamy, że nie ma podstaw do odrzucenia hipotezy H_0 .

I. Test zgodności Pearsona

1. Formułujemy hipotezy:

H_0 : populacja generalna ma rozkład ...

H_1 : populacja generalna nie ma tego rozkładu

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

gdzie r to liczba przedziałów w szeregu, n_i to liczebności empiryczne w próbce, p_i prawdopodobieństwa/odsetki teoretyczne, n liczebność ogólna próbki, np_i liczebności teoretyczne

- Prawdopodobieństwa p_i w rozkładzie normalnym odczytujemy odpowiednio odczytując tablice
- Prawdopodobieństwa p_i w rozkładzie Poissona odczytujemy liczymy ze wzoru:
$$p_i = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}, \text{ gdzie } \lambda \text{ jest średnią rozkładu (najczęściej przyjmujemy tu średnią z próbki } \bar{X} \text{)}$$
- Prawdopodobieństwa p_i w rozkładzie Bernoulliego/dwumianowym liczymy ze wzoru:
$$p_i = \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}, \text{ gdzie } p \text{ jest prawdopodobieństwem sukcesu w pojedynczej próbie}$$

3. Tworzymy i rysujemy prawostronny obszar krytyczny dla rozkładu chi-kwadrat, dla $r - k - 1$ stopni swobody, gdzie k oznacza liczbę parametrów w rozkładzie teoretycznym ($k = 2$ w rozkładzie normalnym, $k = 1$ w rozkładach Poissona i dwumianowym/Bernoulliego)

4. Sprawdzamy, czy statystyka znalazła się w obszarze krytycznym. Jeśli tak – odrzucamy hipotezę H_0 na rzecz hipotezy alternatywnej H_1 . Jeśli nie – stwierdzamy, że nie ma podstaw do odrzucenia hipotezy H_0 .

II. Test losowości próby

II.a Dla małej liczebności próby

1. Formułujemy hipotezy:

H_0 : próba ma charakter losowy

H_1 : populacja generalna nie ma tego rozkładu

2. Porządkujemy próbę w kolejności rosnącej i wszystkim wynikom przyporządkowujemy literę a , jeśli jest on mniejszy od mediany; b , jeśli większy (jeśli równy – pomijamy).

3. Ustawiamy z powrotem wyniki otrzymując ciąg znaków $aababb...$. Liczbę serii oznaczamy przez k . Liczby znaków a i b oznaczamy przez n_1 i n_2 .

4. Z tablic rozkładu serii odczytujemy wartości graniczne k_1 i k_2 takie, żeby $P(k \leq k_1) = \frac{\alpha}{2}$;

$$P(k \leq k_2) = 1 - \frac{\alpha}{2}.$$

5. Sprawdzamy, czy k należy do obszaru krytycznego $k < k_1 \vee k > k_2$ i piszemy odpowiedź.

II.b Dla dużej liczebności próby

1. Formułujemy hipotezy:

H_0 : populacja generalna ma rozkład ...

H_1 : populacja generalna nie ma tego rozkładu

2. Porządkujemy próbę w kolejności rosnącej i wszystkim wynikom przyporządkowujemy literę a , jeśli jest on mniejszy od mediany; b , jeśli większy (jeśli równy – pomijamy).

3. Ustawiamy z powrotem wyniki otrzymując ciąg znaków $aababb...$. Liczbę serii oznaczamy przez k . Liczby znaków a i b oznaczamy przez n_1 i n_2 .

4. Obliczamy statystykę:
$$Z = \frac{k - \left(\frac{2n_1n_2}{n_1 + n_2} + 1 \right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}},$$

5. Tworzymy i obustronny obszar krytyczny dla rozkładu normalnego.

6. Sprawdzamy, czy statystyka znalazła się w obszarze krytycznym. Jeśli tak – odrzucamy hipotezę H_0 na rzecz hipotezy alternatywnej H_1 . Jeśli nie – stwierdzamy, że nie ma podstaw do odrzucenia hipotezy H_0 .

III. Test zgodności dwóch rozkładów

III.a Dla małej liczebności próby

1. Formułujemy hipotezy:

H_0 : próbki pochodzą z tej samej populacji (o tym samym rozkładzie)

H_1 : próbki pochodzą z tej samej populacji

2. Porządkujemy wyniki obu próbek w ciąg o kolejności niemalejącej (w przypadku takich samych wartości w obu próbkach najpierw wypisujemy wyniki z pierwszej, a potem z drugiej) i wszystkim wynikom z pierwszej próbki przyporządkowujemy literę a , a wszystkim wynikom z próbki drugiej literkę b .

3. Liczbę serii oznaczamy przez k . Liczby znaków a i b oznaczamy przez n_1 i n_2 .

4. Z tablic rozkładu serii odczytujemy wartość graniczną k_1 taką, żeby $P(k \leq k_1) = \alpha$ (lewostronny obszar krytyczny).

5. Sprawdzamy, czy k należy do obszaru krytycznego $k \leq k_1$ i piszemy odpowiedź.

III.b Dla dużej liczebności próby

1. Formułujemy hipotezy:

H_0 : próbki pochodzą z tej samej populacji (o tym samym rozkładzie)

H_1 : próbki pochodzą z tej samej populacji

2. Porządkujemy wyniki obu próbek w ciąg o kolejności niemalejącej (w przypadku takich samych wartości w obu próbkach najpierw wypisujemy wyniki z pierwszej, a potem z drugiej) i wszystkim wynikom z pierwszej próbki przyporządkowujemy literę a , a wszystkim wynikom z próbki drugiej literkę b .

3. Obliczamy statystykę:
$$Z = \frac{k - \left(\frac{2n_1n_2}{n_1 + n_2} + 1 \right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}},$$

4. Tworzymy i obustronny obszar krytyczny dla rozkładu normalnego.

5. Sprawdzamy, czy statystyka znalazła się w obszarze krytycznym. Jeśli tak – odrzucamy hipotezę H_0 na rzecz hipotezy alternatywnej H_1 . Jeśli nie – stwierdzamy, że nie ma podstaw do odrzucenia hipotezy H_0 .