

TESTY NORMALNOŚCI

Test zgodności χ^2

Hipoteza zerowa H_0 (Cecha X populacji ma rozkład normalny).

Hipoteza alternatywna H_1 (Cecha X populacji nie ma rozkładu normalnego).

Weryfikacja powyższych hipotez za pomocą tzw. testu χ^2 przebiega następująco:

1. Pobieramy liczną próbę ($n > 80$).
Prezentujemy ją w szeregu rozdzielczym klasowym w r klasach.
2. Obliczamy na podstawie próby estymatory największej wiarygodności nieznanych parametrów m i σ (średnia i odchylenie standardowe z próby).
3. Przyjmujemy, że cecha X ma rozkład normalny.

4. Dla każdego przedziału klasowego

$A_i = [a_i; a_{i+1})$ obliczamy

prawdopodobieństwo

$$p_i = P(X \in A_i) = P(a_i \leq X < a_{i+1}) = F(a_{i+1}) - F(a_i)$$

5. Obliczamy

$$u_n = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^r \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$$

gdzie n_i jest liczebnością (empiryczną) klasy A_i .

$\hat{n}_i = np_i$ jest liczebnością teoretyczną klasy A_i

6. Wyznaczamy zbiór krytyczny
prawostronny $K = (k; \infty)$,
gdzie k wyznaczamy z tablicy rozkładu χ^2
dla $r - 3$ stopni swobody
i dla prawdopodobieństwa α (równemu
poziomowi istotności).
7. Podejmujemy decyzję:
odrzucamy hipotezę H_0 , gdy $u_n \in K$
przyjmujemy hipotezę H_0 , gdy $u_n \notin K$

Uwaga. Do obliczania prawdopodobieństw p_i , pierwsza i ostatnia klasa szeregu rozdzielczego powinny mieć postać $A_1 = (-\infty ; a_2)$, $A_r = (a_r ; \infty)$ i do każdej z nich powinno należeć co najmniej 5 elementów próby. Do pozostałych klas powinno należeć co najmniej 10 elementów próby. Klas nie może być mniej niż 4.

Przykład

Pobrano próbę 200 elementową. Otrzymane wyniki prezentowane są w poniższym szeregu rozdzielczym przedziałowym

Nr klasy i	Klasa $< a_i; a_{i+1})$	Liczebność n_i
1	$<600; 800)$	2
2	$<800; 1\ 000)$	10
3	$<1\ 000; 1\ 200)$	20
4	$<1\ 200; 1\ 400)$	30
5	$<1\ 400; 1\ 600)$	56
6	$<1\ 600; 1\ 800)$	42
7	$<1\ 800; 2\ 000)$	21
8	$<2\ 000; 2\ 200)$	13
9	$<2\ 200; 2\ 400)$	5
10	$<2\ 400; 2\ 600)$	1
	Suma	200

Na poziomie istotności $\alpha = 0,05$ sprawdź hipotezy:

H_0 (Cecha X populacji ma rozkład normalny) i

H_1 (Cecha X populacji nie ma rozkładu normalnego).

Rozwiązanie

Nieznane parametry to m i σ ($l = 2$) ich estymatorami są $\bar{x}$ i s .

Obliczenia $\bar{x}$ i s

Nr klasy i	Klasa $< a_i; a_{i+1})$	Liczebność n_i	Środek klasy w_i	$w_i n_i$	$n_i (w_i - \bar{x})^2$
1	<600; 800)	2	700	1 400	1 411 200
2	<800; 1 000)	10	900	9 000	4 096 000
3	<1 000; 1 200)	20	1 100	22 000	3 872 000
4	<1 200; 1 400)	30	1 300	39 000	1 728 000
5	<1 400; 1 600)	56	1 500	84 000	89 600
6	<1 600; 1 800)	42	1 700	71 400	1 075 200
7	<1 800; 2 000)	21	1 900	39 900	2 721 600
8	<2000; 2200)	13	2100	27 300	4 076 800
9	<2200; 2400)	5	2300	11 500	2 888 000
10	<2400; 2600)	1	2500	2 500	921 600
	Suma	200		308 000	22 880 000

$$\bar{x} = \frac{308000}{200} = 1540 \text{ [zł]},$$

$$s^2 = \frac{22880000}{200} = 114400 \text{ [zł]}^2,$$

$$s = \sqrt{114400} = 338,2 \text{ [zł]}$$

Obliczenia u_{100}

i	$< a_i; a_{i+1})$	n_i	a_i	a_{i+1}	$\frac{a_i - \bar{x}}{s}$	$\frac{a_{i+1} - \bar{x}}{s}$	$\Phi(\frac{a_i - \bar{x}}{s})$	$\Phi(\frac{a_{i+1} - \bar{x}}{s})$	p_i	$n p_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty; 1000)$	12	$-\infty$	1000	$-\infty$	-1,60	0	0,0552	0,05517	11,03	0,084659
2	$<1000; 1200)$	20	1000	1200	-1,60	-1,01	0,05517	0,1574	0,10220	20,44	0,009499
3	$<1200; 1400)$	30	1200	1400	-1,01	-0,41	0,15737	0,3395	0,18208	36,42	1,130557
4	$<1400; 1600)$	56	1400	1600	-0,41	0,18	0,33945	0,5704	0,23095	46,19	2,083142
5	$<1600; 1800)$	42	1600	1800	0,18	0,77	0,57041	0,7790	0,20858	41,72	0,001933
6	$<1800; 2000)$	21	1800	2000	0,77	1,36	0,77899	0,9131	0,13412	26,82	1,264544
7	$<2000; \infty)$	19	2000	∞	1,36	∞	0,91311	1	0,08689	17,38	0,151291
								Suma	1,00000	200,00	4,730000

$$u_{100} = 4,73.$$

Wyznaczamy zbiór krytyczny prawostronny $K = < k; \infty$.

Liczbę k odczytujemy z tablicy rozkładu χ^2 dla $r - l - 1 = 7 - 2 - 1 = 4$ stopni swobody i prawdopodobieństwa $\alpha = 0,05$.

Mamy $k = 9,488$, więc $K = < 9,488; \infty$.

Ponieważ $u_{100} = 4,73 \notin K$, więc hipotezę, że cecha ma rozkład normalny przyjmujemy.

Hipotezę tę można dopiero odrzucić na poziomie istotności 0,32, gdyż zbiór krytyczny $K = < 4,73; \infty$ otrzymujemy właśnie na tym poziomie.

Test Shapiro-Wilka

Wysuwamy hipotezy

H_0 (Cecha X populacji ma rozkład normalny)

H_1 (Cecha X populacji nie ma rozkładu normalnego)

Zweryfikujemy powyższe hipotezy na podstawie małej próby, na poziomie istotności α , za pomocą testu Shapiro-Wilka.

1. Porządkujemy próbę niemalejąco

$$(x_{(1)}, x_{(2)}, \dots, x_{(n)})$$

2. Obliczamy statystykę

$$U_n = \frac{\left[\sum_{i=1}^{[n/2]} a_{n-i+1} (X_{(n-i+1)} - X_{(i)}) \right]^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

gdzie: $[n/2]$ jest częścią całkowitą liczby $n/2$, a współczynniki $a_1, a_2, \dots, a_{n/2}$, odczytujemy z odpowiedniej tablicy :

$\begin{smallmatrix} i \\ n \end{smallmatrix}$	1	2	3	4	5	6	7	8	9	10
8	0,6052	0,3164	0,1743	0,0561	—	—	—	—	—	—
10	0,5739	0,3291	0,2141	0,1224	0,0399	—	—	—	—	—
12	0,5475	0,3325	0,2347	0,1586	0,0922	0,0303	—	—	—	—
14	0,5251	0,3318	0,2460	0,1802	0,1240	0,0727	0,0240	—	—	—
15	0,5150	0,3306	0,2495	0,1878	0,1353	0,0880	0,0433	0	—	—
16	0,5056	0,3290	0,2521	0,1939	0,1447	0,1005	0,0593	0,0196	—	—
18	0,4886	0,3253	0,2553	0,2027	0,1587	0,1197	0,0837	0,0496	0,0163	—
20	0,4734	0,3211	0,2565	0,2085	0,1686	0,1334	0,1013	0,0711	0,0422	0,0140

3. Rozpatrujemy zbiór krytyczny: $K = < 0; k >$

gdzie k odczytujemy dla poziomu istotności α i danego n z tablicy testu Shapiro-Wilka:

(tablica testu Shapiro-Wilka dla $\alpha = 0,05$)

n	8	10	12	14	15	16	18	20
k	0,818	0,842	0,859	0,874	0,881	0,887	0,897	0,905

4. Decyzje:

Jeśli $u_n \in K$ to H_0 odrzucamy.

Jeśli $u_n \notin K$ to nie ma podstaw do odrzucenia H_0 .

Przykład

Dana jest uporządkowana próba 18 elementowa: 124, 142, 149, 156, 161, 168, 173, 179, 182, 193, 197, 204, 219, 228, 237, 252, 259, 274. Na poziomie istotności 0,05 sprawdzić testem Shapiro-Wilka hipotezę o normalności rozkładu badanej cechy.

Rozwiązanie

Średnia wynosi 194,3.

Suma kwadratów odchyłeń od średniej $\sum_{i=1}^n (x_i - \bar{x})^2 = 31375,6$.

$$u_n = \frac{[0,4886(274 - 124) + 0,3253(259 - 142) + \dots + 0,0163(193 - 182)]^2}{31375,6} = 0,97$$

$K = \langle 0; 0,897 \rangle$, zatem $u_n \notin K$ i hipotezę o normalności rozkładu badanej cechy należy przyjąć.

Test normalności (test Kołmogorowa – Lillieforsa)

Wysuwamy dwie hipotezy:

H_0 – X ma rozkład normalny,

H_1 – X nie ma rozkładu normalnego.

Pobieramy próbę co najmniej 30 elementową. Na podstawie próby wyznaczamy realizację estymatorów parametrów m i σ czyli średnią i odchylenie standardowe z próby.

Obliczamy wartość statystyki

$$U_n = \sup_x |F_n(x) - F(x)|$$

Gdzie $F_n(x)$ jest dystrybuantą empiryczną, $F(x)$ dystrybuantą rozkładu normalnego $N(\bar{x}, s)$.

Wyznaczamy zbiór krytyczny $K = < k, \infty)$, gdzie k odczytujemy z tablicy dla danego poziomu istotności.

Krytyczne wartości testu Kołmogorowa – Lillieforsa dla $\alpha = 0,05$:

n – liczebność próby, k – lewy koniec zbioru krytycznego.

n	35	40	45	50	55	60	65
k	0,1498	0,1401	0,1321	0,1253	0,1193	0,1144	0,1099

n	70	75	80	85	90	95	100
k	0,1059	0,1023	0,0991	0,0961	0,0934	0,0909	0,0886

Podajemy decyzję:

- odrzucamy hipotezę H_0 , gdy $u_n \in K$
- przyjmujemy hipotezę H_0 , gdy $u_n \notin K$