Kryterium (test) zgodności dla rozkładu jednostajnego w (0, 1).

Uwaga.

Jeśli x_1, x_2, …., x_n jest próbą dla rozkładu o dystrybuancie F, to wielkości

$$y_1=F(x_1),\ y_2=F(x_2),\ \ldots,\ y_n=F(x_n)$$

mają rozkład jednostajny w [0, 1].

Stąd wynika doniosłość testów zgodności dla rozkładu jednostajnego.

Test Watsona.

H_0 (X ma rozkład jednostajny w [0, 1].

α - poziom istotności.

Statystyka:

$$U_n^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 - \frac{2}{n}\sum_{i=1}^{n} ix_i + (n+1)\overline{x} + \frac{n}{12}$$

x_i - uporządkowane dane niemalejąco z przedziału [0, 1].

Oraz przyjmujemy $x_0 = 0$.

Zbiór krytyczny K =<k, ∞).

Dla n > 10 i statystyki

$$W_n^2 = \left(U_n^2 - \frac{0,1}{n} + \frac{0,1}{n^2}\right)\left(1 + \frac{0,8}{n}\right)$$

wartości krytyczne k odczytujemy z tablicy:

	α = 0,1	α = 0,05	α = 0,01
k	0,152	0,187	0,267

Dla statystyki U_n^2 można też wyznaczyć wartość krytyczną k na podstawie odczytu z tablicy rozkładu chi kwadrat:

$$k = \frac{21n - 56}{840\left(n - \frac{3}{2}\right)} + \frac{1}{42n}\left(n - \frac{3}{2}\right)k'$$

gdzie k' odczytujemy z tablicy rozkładu chi kwadrat dla ustalonego poziomu istotności i $\dfrac{49n(n-1)}{20\left(n - \frac{3}{2}\right)^2}$ stopni swobody (po zaokrągleniu do wartości całkowitej).

Test Shermana.

Statystyka:

$$U = \frac{1}{2}\sum_{i=1}^{n}\left| x_i - x_{i-1} - \frac{1}{n+1} \right|$$

x_i - dane uporządkowane niemalejąco,

Oraz przyjmujemy $x_0 = 0$.

Zbiór krytyczny K =<k, ∞).

Wartości krytyczne k odczytujemy z tablicy:

n	α = 0,1	α = 0,05	α = 0,01
1	0,450	0,475	0,495
2	0,484	0,537	0,609
3	0,467	0,518	0,614
4	0,468	0,509	0,589
5	0,462	0,502	0,574
6	0,458	0,494	0,562
7	0,454	0,488	0,551

8	0,451	0,482	0,542
9	0,448	0,477	0,534
10	0,445	0,473	0,527
11	0,442	0,469	0,521
12	0,440	0,466	0,516
13	0,438	0,463	0,511
14	0,436	0,460	0,506
15	0,434	0,458	0,502
16	0,433	0,455	0,498
17	0,431	0,453	0,495
18	0,430	0,451	0,491
19	0,429	0,449	0,489
20	0,427	0,448	0,486

I. Dla n > 20 powyższa statystyka ma rozkład asymptotycznie normalny o parametrach:

$$m = \left(\frac{n}{n+1} \right)^{n+1} \approx \frac{1}{e} = 0,36788$$

$$\sigma^2 = \frac{2n^{n+2} + n(n-1)^{n+2}}{(n+2)(n+1)^{n+2}} - \left(\frac{n}{n+1}\right)^{2n+2} \approx \frac{2e-5}{e^2} \cdot \frac{1}{n} = \frac{0{,}05908}{n}$$

II. Dla n > 20 powyższą statystykę można aproksymować rozkładem normalnym N(0, 1):

$$\widetilde{U} = W - \frac{0{,}0955}{\sqrt{n}}\left(W^2 - 1\right)$$

Gdzie

$$W = \frac{U - 0{,}3679}{0{,}2431\left(1 - \frac{0{,}605}{n}\right)}\left(1 - \frac{1}{2n}\right)\sqrt{n}$$

Test Morana.

Statystyka:

$$U = -\sum_{i=1}^{n+1} \ln\big((n+1)d_i\big)$$

x_i - dane uporządkowane niemalejąco,

$$d_i = x_i - x_{i-1}, \quad d_1 = x_1, \qquad d_{n+1} = 1 - x_n,$$

Zbiór krytyczny K =(-∞, k_1> U <k_2, ∞).

Wartości krytyczne k odczytujemy z tablicy:

n	α = 0,1 k_1	α = 0,1 k_2	α = 0,05 k_1	α = 0,05 k_2	α = 0,01 k_1	α = 0,01 k_2
2	0,130	2,756	0,064	3,554	0,012	5,370
3	0,356	3,699	0,215	4,592	0,070	6,579
4	0,643	4,573	0,430	5,546	0,180	7,675
5	0,967	5,408	0,689	6,449	0,334	8,705
6	1,317	6,214	0,979	7,319	0,523	9,688
7	1,687	7,001	1,292	8,162	0,741	10,637
8	2,072	7,772	1,625	8,987	0,982	11,559

9	2,469	8,531	1,973	9,795	1,242	12,460
10	2,876	9,279	2,333	10,591	1,519	13,343
11	3,292	10,018	2,704	11,375	1,810	14,210
12	3,715	10,750	3,085	12,150	2,113	15,064
13	4,145	11,475	3,479	12,916	2,428	15,907
14	4,581	12,194	3,870	13,675	2,752	16,740
15	5,021	12,909	4,272	14,428	3,085	17,563
16	5,466	13,618	4,681	15,174	3,425	18,378
17	5,916	14,324	5,094	15,916	3,774	19,186
18	6,369	15,025	5,513	16,652	4,128	19,987
19	6,825	15,721	5,936	17,384	4,489	20,782
20	7,285	16,418	6,362	18,112	4,855	21,571

Dane:

0,047; 0,05; 0,15; 0,18; 0,29; 0,48; 0,52; 0,61; 0,72; 0,91.

Uwaga

Statystyka

$$W = \frac{12(n+1)}{7n+8}U$$

Ma rozkład chi kwadrat o *n* stopniach swobody.

Uwaga

Statystyka

$$Z = \frac{\left(\frac{12(n+1)}{n(7n+8)}U\right)^{\frac{1}{3}} - \left(1 - \frac{2}{9n}\right)}{\sqrt{\frac{2}{9n}}}$$

Ma rozkład N(0, 1).

Test zgodności χ^2

Hipoteza zerowa $H_0($Cecha X populacji ma rozkład jednostajny w przedziale [a, b]).

Hipoteza alternatywna $H_1($Cecha *X* populacji nie ma rozkładu jednostajnego w przedziale [a, b]).

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. Przyjmujemy, że cecha *X* ma rozkład jednostajny w przedziale [a, b].
3. Dla każdego przedziału klasowego $A_i =< a_i ; a_{i+1})$ obliczamy prawdopodobieństwo

$$p_i = P(X \in A_i) = (b-a)/r$$

4. 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

5. Wyznaczamy zbiór krytyczny prawostronny $K =< k\ ;\ \infty)$, gdzie *k* wyznaczamy z tablicy rozkładu χ^2 dla r –1 stopniami swobody i dla prawdopodobieństwa α (równemu poziomowi istotności).

6. Podejmujemy decyzję:

odrzucamy hipotezę H_0, gdy $u_n \in K$

przyjmujemy hipotezę H_0, gdy $u_n \notin K$

Uwaga. Do pierwszej i ostatniej klasy szeregu rozdzielczego powinno należeć co najmniej 5 elementów próby. Do pozostałych klas powinno należeć co najmniej 10 elementów próby. Klas nie może być mniej niż 4.