Testy zgodności

H_0 : Cecha X ma rozkład F

F jest dowolnym rozkładem prawdopodobieństwa

Test chi–kwadrat zgodności

F jest rozkładem ciągłym

Test Kołmogorowa

F jest rozkładem normalnym

Test Shapiro–Wilka

Test chi–kwadrat zgodności

Klasa	Liczebność
1	n_1
2	n_2
$\vdots$	$\vdots$
k	n_k

$$\chi^2_{\text{emp}} = \sum_{i=1}^{k} \frac{(n_i - n_i^t)^2}{n_i^t}$$

$$n_i^t = N p_i^t, \ N = \sum_{i=1}^{k} n_i,$$

$$p_i^t = P_F\{X \text{ przyjęła wartość z klasy } i\}$$

Wartość krytyczna $\chi^2(\alpha; k-u-1)$ (u jest liczbą nieznanych parametrów hipotetycznego rozkładu F)

Wniosek. Jeżeli $\chi^2_{\text{emp}} > \chi^2(\alpha; k-u-1)$, to hipotezę H_0 odrzucamy

Przykład . Na podstawie poniższej próby zbadać, czy obserwowana cecha X ma rozkład normalny.

Klasa		Liczebność
$-\infty$	x_1	n_1
x_1	x_2	n_2
$\vdots$	$\vdots$	$\vdots$
x_{k-2}	x_{k-1}	n_{k-1}
x_{k-1}	∞	n_k
		N

$$H_0 : \text{Cecha } X \text{ ma rozkład normalny } N(\mu, \sigma^2)$$

Wyznaczenie wartości statystyki χ^2_{emp}
Wyznaczenie prawdopodobieństw teoretycznych

$$p_i^t = P\{x_i < X < x_{i+1}\} = \Phi\left(\frac{x_{i+1} - \mu}{\sigma}\right) - \Phi\left(\frac{x_i - \mu}{\sigma}\right)$$

Z próby wyznaczamy $\bar{x}$ oraz s^2

$$p_i^t = \Phi\left(\frac{x_{i+1} - \bar{x}}{s}\right) - \Phi\left(\frac{x_i - \bar{x}}{s}\right) = \Phi(z_{i+1}) - \Phi(z_i)$$

Klasa		n_i	p_i^t	n_i^t	$(n_i - n_i^t)^2/n_i^t$
$-\infty$	x_1	n_1	p_1^t	$n_1^t = Np_1^t$	$(n_1 - n_1^t)^2/n_1^t$
x_1	x_2	n_2	p_2^t	$n_2^t = Np_2^t$	$(n_2 - n_2^t)^2/n_2^t$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{k-2}	x_{k-1}	n_{k-1}	p_{k-1}^t	$n_{k-1}^t = Np_{k-1}^t$	$(n_{k-1} - n_{k-1}^t)^2/n_{k-1}^t$
x_{k-1}	∞	n_k	p_k^t	$n_k^t = Np_k^t$	$(n_k - n_k^t)^2/n_k^t$
		N	1	N	χ^2_{emp}

Wartość krytyczna $\chi^2(\alpha; k-2-1)$

. . . .

Test Kołmogorowa

Próba $X_1, \ldots, X_n$

Próbę $X_1, \ldots, X_n$ porządkujemy niemalejąco

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

Tak uporządkowane obserwacje nazywamy statystykami pozycyjnymi

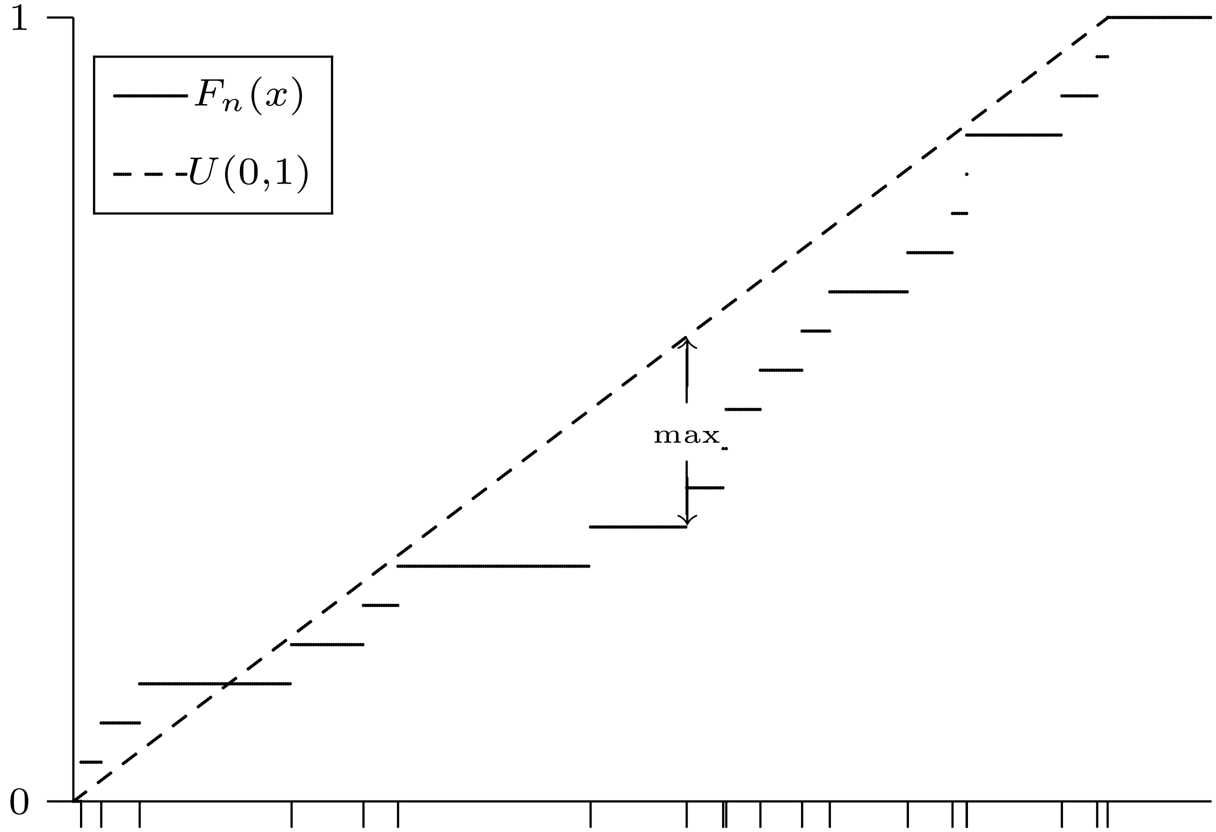
Statystyka testowa

$$D_n = \max_{1 \leq i \leq n} \left\{ \max \left\{ \left| F(X_{(i)}) - \frac{i-1}{n} \right|, \left| \frac{i}{n} - F(X_{(i)}) \right| \right\} \right\}$$

Wartość krytyczna testu Kołmogorowa $D(\alpha; n)$

Jeżeli $D_n > D(\alpha; n)$, to hipotezę H_0 odrzucamy

$$H_0 : X \sim U(0,1)$$



$$H_0 : \text{Cecha } X \text{ ma rozkład normalny}$$

Test Shapiro–Wilka

Próbę $X_1, \ldots, X_n$ porządkujemy niemalejąco

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

Tak uporządkowane obserwacje nazywamy statystykami pozycyjnymi

Statystyka testowa

$$W = \frac{\left(\sum_{i=1}^{[n/2]} a_{i:n}(X_{(n-i+1)} - X_{(i)})\right)^2}{\mathrm{var}X}$$

$a_{i:n}$ są stablicowanymi współczynnikami

$$[n/2] = \begin{cases} n/2, & \text{dla } n \text{ parzystych} \\ (n-1)/2, & \text{dla } n \text{ nieparzystych} \end{cases}$$

Wartość krytyczna testu Shapiro–Wilka $W_n(\alpha)$

Jeżeli $W \leq W_n(\alpha)$, to hipotezę H_0 odrzucamy.

Przykład . Z cechy X pobrano próbę ($n = 19$): 12.4, 14.2, 14.9, 15.6, 16.1, 17.3, 17.9, 18.2, 18.6, 19.3, 19.7, 20.4, 21.9, 22.8, 23.7, 25.2, 25.9, 27.4.

$$H_0 : \text{Cecha } X \text{ ma rozkład normalny}$$

Test Shapiro–Wilka ($\alpha = 0.05$)

$$\bar{x} = 19.3842,\ \text{var}x = 730.57$$

Licznik statystyki W

i	$x_{(19-i+1)} - x_{(i)}$	$a_{i:19}$	$a_{i:19}(x_{(19-i+1)} - x_{(i)})$
1	27.4−12.4=15.0	0.4808	7.21200
2	25.9−14.2=11.7	0.3232	3.78144
3	25.2−14.9=10.3	0.2561	2.63783
4	23.7−15.6= 8.1	0.2059	1.66779
5	22.8−16.1= 6.7	0.1641	1.09947
6	21.9−16.8= 5.1	0.1271	0.64821
7	20.4−17.3= 3.1	0.0932	0.28892
8	19.7−17.9= 1.8	0.0612	0.11016
9	19.3−18.2= 1.1	0.0303	0.03333
			17.47915

Statystyka testowa $W = \frac{305.52}{730.57} = 0.418$
Wartość krytyczna $W_{19}(0.05) = 0.901$
Ponieważ $W < W_{19}(0.05)$, więc hipotezę odrzucamy

. . . .