6. TESTOWANIE HIPOTEZ NIEPARAMETRYCZNYCH

Hipotezy nieparametryczne są np. opiniami dotyczącymi typu rozkładu, przypuszczeniami o jednakowym rozkładzie dwóch populacji czy sądami o korelacji zjawisk. W zależności od typu weryfikowanej hipotezy statystycznej wyróżnia się np:

- testy zgodności – w których stawiana jest hipoteza o zgodności próby z pewnym rozkładem teoretycznym (np. normalnym), najczęściej stosowanymi testami zgodności są: test χ^2 oraz test Kołmogorowa – Smirnowa,
- testy niezależności – w których hipoteza zerowa jest sądem o niezależności dwóch cech a hipoteza alternatywna o ich zależności.

6.1. Test zgodności χ^2

W teście χ^2 dla zweryfikowania hipotezy o zgodności rozkładu próby z pewnym rozkładem teoretycznym porównuje się liczebności szeregu empirycznego (próby) z liczebnościami szeregu teoretycznego. Do wyznaczenia wartości statystyki testowej niezbędne jest pogrupowanie wyników próby w *szereg rozdzielczy*. Następnie wyznaczana jest wartość statystyki:

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i},$$

gdzie: r – liczba klas szeregu rozdzielczego; n_i – liczebność i–tej klasy; n – liczebność próby; p_i – prawdopodobieństwo teoretyczne, że zmienna losowa o weryfikowanym typie rozkładu przyjmie wartość należącą do i–tej klasy.

Przy założeniu prawdziwości hipotezy H_0 (rozkład z próby jest zgodny z rozkładem teoretycznym) statystyka χ^2 ma rozkład χ^2 o $v = r - k - 1$ stopniach swobody (k – liczba szacowanych parametrów rozkładu). Obszar krytyczny w teście budowany jest jako prawostronny.

Uwagi: Liczebność przedziałów szeregu nie powinna być mniejsza od 5. Końce przedziałów pierwszego i ostatniego przyjmuje się w nieskończoności.

Przykład 1.

Wykonano 100 pomiarów długości detalu. Średnia długość wyniosła $\bar{x} \approx 20.96$ a odchylenia standardowe $s \approx 0.69$. Dane zebrano w postaci szeregu rozdzielczego. Zweryfikować na poziomie istotności $\alpha = 0.01$ hipotezę, że rozkład długości jest rozkładem normalnym.

Długość	[19, 19.5]	[19.5 20]	[20 20.5]	[20.5 21]	[21 21.5]	[21.5 22]	[22 22.5]	[22.5 23]
Liczność	1	6	18	29	26	12	6	2

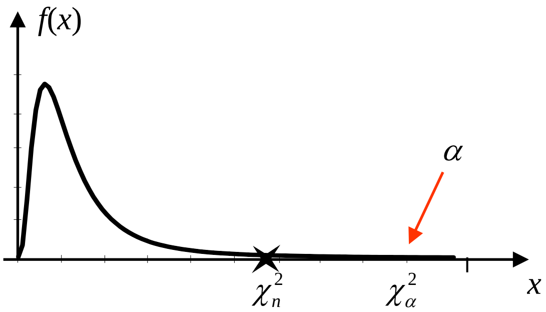
Obliczenia wartości statystyki χ^2 wygodnie jest przedstawiać w formie tabelarycznej. Ze względu na to, że liczebności w dwóch przedziałach były mniejsze od 5, przedziały te zostały połączone z sąsiednimi.

Lp	**Odchylenie od nominalnej średnicy**		$\boldsymbol{n_i}$	$\boldsymbol{p_i}$	$\boldsymbol{n*p_i}$	$\boldsymbol{n_i-n*p_i}$	$\boldsymbol{(n_i-n*p_i)^2}$	$\boldsymbol{\frac{(n_i-n*p_i)^2}{n*p_i}}$
	od	**do**						
1		20	7	0.0828	8.28	-1.28	1.639	0.1962
2	20	20,5	18	0.1699	16.99	1.01	1.017	0.0599
3	20,5	21	29	0.2690	26.90	2.10	4.399	0.1635
4	21	21,5	26	0.2591	25.91	0.09	0.008	0.0003
5	21,5	22	12	0.1518	15.18	-3.18	10.095	0.6651
6	22		8	0.0674	6.74	1.26	1.590	0.2360
Σ	100			1			χ^2	1.3228

Obliczona na podstawie wyników z próby wartość statystyki testowej wyniosła więc:

$$\chi_n^2 \approx 1.3228.$$

Obszar krytyczny wyznacza się wykorzystując rozkład χ^2 o 3 stopniach swobody (statystyka χ^2 została zbudowana na podstawie r = 6 klas szeregu rozdzielczego danych, 2 parametry rozkładu były szacowane)



$$\chi_\alpha^2 = F^{-1}_{\chi^2(6-2-1)}(1-\alpha) = F^{-1}_{\chi^2(3)}(0.99) \approx 11.35$$

Wartość statystyki testowej leży poza obszarem krytycznym, nie można odrzucić hipotezy H_0.

Graniczny poziom istotności p–value dla testu prawostronnego wynosi:

$$p\text{–}value = 1 - F_{\chi^2(3)}\left(\chi_n^2\right) = 1 - F_{\chi^2(5)}(1.3228) \approx 0.72$$

Założony poziom istotności α jest niższy od poziomu granicznego nie można więc odrzucić hipotezy H_0.

6.2. Test zgodności λ Kołmogorowa (Kołmogorowa – Smirnowa)

W teście λ dla zweryfikowania hipotezy o zgodności rozkładu próby z pewnym rozkładem teoretycznym porównuje się dystrybuanty empiryczną i teoretyczną i jeśli obydwie dystrybuanty mają we wszystkich badanych punktach zbliżone wartości to uznaje się, że hipoteza o zgodności rozkładu próby z badanym rozkładem teoretycznym nie może być odrzucona. W przypadku testu Kołmogorowa zakłada się, że dystrybuanta teoretyczna jest ciągła – test ten nie może być więc stosowany do zbadania zgodności rozkładu z próby z rozkładem skokowym. Ograniczenia tego nie ma omówiony powyżej test χ^2.

Do budowy statystyki testowej λ Kołmogorowa wykorzystywana jest największa różnica pomiędzy dystrybuantami:

$$\lambda = D\sqrt{n}\ ,$$

gdzie: D – maksymalna różnica pomiędzy dystrybuantami empiryczną i teoretyczną, definiowana jako: $D = \sup_x |F_n(x) - F(x)|$, sup – supremum, kres górny zbioru, $F_n(x)$ – dystrybuanta empiryczna (wyznaczana na podstawie rozkładu empirycznego przypisującego każdej wartości z próby prawdopodobieństwo $1/n$), $F(x)$ – dystrybuanta teoretyczna.

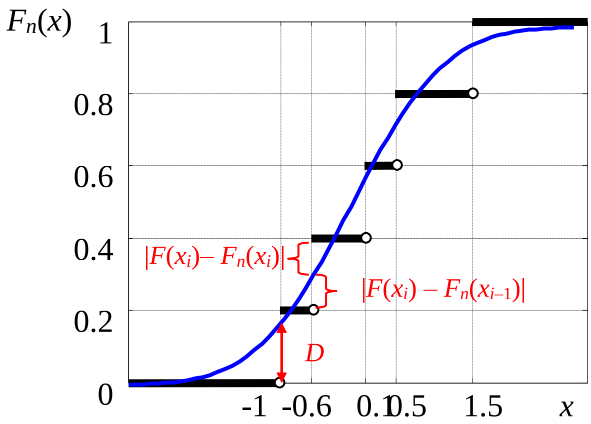
Statystyka λ przy prawdziwości hipotezy o zgodności rozkładu z wybranym rozkładem teoretycznym ma rozkład λ Kołmogorowa, obszar krytyczny w teście budowany jest jako prawostronny.

Przykład 2

Wykonano 5 pomiarów długości detalu: -1.0, 0.1, -0.6, 0.5, 1.5. Zweryfikować na poziomie istotności $\alpha = 0.01$ hipotezę, że rozkład długości jest rozkładem normalnym standaryzowanym.

Dystrybuantę empiryczną wyznacza się przypisując każdej wartości z próby prawdopodobieństwo $p = 1/n$, czyli w tym przypadku $p = 1/5 = 0.2$. Porządkując wyniki pomiarów dystrybuantę empiryczną można przedstawić w postaci tabelarycznej oraz w postaci wykresu:

x_i	-1.0	-0.6	0.1	0.5	1.5
p_i	0.2	0.2	0.2	0.2	0.2
F_n	0.2	0.4	0.6	0.8	1.0

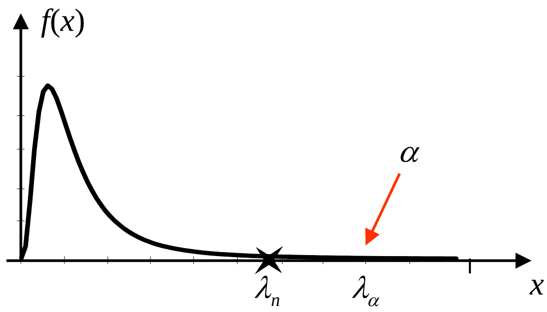


Na powyższym wykresie, dodatkowo linią w kolorze niebieskim, zaznaczono dystrybuantę standaryzowanego rozkładu normalnego. W teście Kołmogorowa wyznaczana jest największa różnica pomiędzy dystrybuantami empiryczną a teoretyczną. Odległości pomiędzy dystrybuantami wyznaczane są dla wszystkich danych z próby. Dla każdego punktu liczone są właściwie dwie odległości: odległość pomiędzy dystrybuantą teoretyczną $F(x_i)$ a dystrybuantą empiryczną $F_n(x_i)$ oraz odległość pomiędzy $F(x_i)$ a $F_n(x_{i-1})$. Obliczenia wartości statystyki λ wygodnie jest przedstawiać w formie tabelarycznej.

x_i	$F(x_i) = \Phi(x_i)$	$F_n(x_i)$	$\lvert F(x_i) - F_n(x_i)\rvert$	$\lvert F(x_i) - F_n(x_{i-1})\rvert$
-1.0	0.159	0.2	0.041	**0.159**
-0.6	0.274	0.4	0.125	0.074
0.1	0.540	0.6	0.060	0.140
0.5	0.692	0.8	0.109	0.092
1.5	0.933	1.0	0.067	0.133

Największa odległość pomiędzy dystrybuantami wystąpiła dla pierwszego punktu próby i wyniosła D = 0.159. Wartość statystyki testowej wynosi więc $\lambda_n = D\sqrt{n} = 0.159\sqrt{5} \approx 0.355$.

Obszar krytyczny wyznacza się wykorzystując odwrotność dystrybuanty rozkładu testowej λ Kołmogorowa:



$$\lambda_\alpha = F_\lambda^{-1}(1-\alpha) = F_\lambda^{-1}(0.99) \approx 1.628$$

Wartość statystyki testowej leży poza obszarem krytycznym, nie można odrzucić hipotezy H_0.

Graniczny poziom istotności p–value dla testu prawostronnego wynosi:

$$p\text{–}value = 1 - F_\lambda(\lambda_n) = 1 - F_\lambda(0.355) \approx 0.999$$

Założony poziom istotności α jest niższy od poziomu granicznego nie można więc odrzucić hipotezy H_0.

6.3. Test niezależności χ^2

Test niezależności wykorzystywany jest do sprawdzenia czy wybrane dwie cechy populacji generalnej są ze sobą związane tzn. czy są zależne.

Dwie zmienne losowe X i Y są niezależne, gdy dla każdej wartości a i b zachodzi równość:

$$P(x \le a)P(y \le b) = P((x \le a) \wedge (y \le b)),$$

zmienne są zależne jeśli istnieje taka para a, b dla której zależność powyższa nie zachodzi.

Do oceny zależności cech mierzalnych wykorzystywana jest analiza regresji i korelacji. Jeśli jedna z cech jest cechą niemierzalną analiza ta nie może być stosowana, może być natomiast zastosowany test niezależności χ^2. Warunkiem stosowalności testu jest duża liczebność próby. W teście niezależności χ^2 wykorzystywana jest ta sama statystyka testowa co w teście zgodności χ^2 – przy założeniu prawdziwości hipotezy o niezależności cech statystyka ta ma rozkład χ^2 o $(r-1)(c-1)$ stopniach swobody (gdzie r i c to liczby grup na które zostały podzielone wartości odpowiednio cechy X i cechy Y). Obszar krytyczny w teście budowany jest jako prawostronny. Do wyznaczenia wartości statystyki χ^2 wykorzystywana jest tzw. *tablica niezależności*. Sposób konstrukcji tablicy zostanie wyjaśniony na poniższym przykładzie.

Przykład 3.

Pewien produkt może być wytwarzany dwiema różnymi metodami. W celu sprawdzenia czy trwałość zależy od użytej metody produkcji przeprowadzono badania dwóch partii produktów. Ocenę trwałości przeprowadzano poprzez wyznaczenie liczby dni pracy do chwili uszkodzenia produktu, wyniki oceny zostały zebrane w poniższej tabeli. Zweryfikować na poziomie istotności α = 0.01 hipotezę, że trwałość produktu zależy od użytej metody produkcji.

Trwałość w dniach	Metoda 1	Metoda 2
40-50	10	10
50-60	30	110
60-70	20	20

Do sprawdzenia niezależności cechy X (trwałość produktu) od cechy Y (metoda produkcji) wykorzystywana jest *tablica niezależności*. Tablica ta ma tyle wierszy ile jest grup wartości cechy X i tyle kolumn ile jest grup wartości cechy Y. W komórkach tablicy wpisywane są liczebności empiryczne n_{ij} określające liczbę elementów próby posiadający jednocześnie cechę X na poziomie wynikającym z opisu i-tego wiersza tablicy i cechę Y wynikającą z opisu j-tej kolumny tablicy. Przedstawiona powyżej tablica jest tzw. *macierzą liczebności empirycznych*. *Tablicę niezależności* otrzymuje się po nałożeniu na macierz liczebności empirycznych *macierzy liczebności teoretycznych*. Do wyznaczenie liczebności teoretycznych konieczne jest wcześniejsze policzenie teoretycznego prawdopodobieństwa p_{ij} wystąpienia wyników posiadających cechę X na poziomie i oraz jednocześnie cechę Y na poziomie j. Prawdopodobieństwa teoretyczne obliczane są przy założeniu, że prawdziwa jest hipoteza H_0 o niezależności cech X i Y, tzn.:

$$p_{ij} = p_i p_j ,$$

gdzie: p_i, p_j – prawdopodobieństwo wystąpienia wyników posiadających odpowiednio: cechę X na poziomie i oraz cechę Y na poziomie j. Prawdopodobieństwa p_i i p_j obliczane są z zależności:

$$p_i = \frac{n_i}{n}; \qquad p_j = \frac{n_j}{n}$$

gdzie: n_i, n_j – liczba wyników posiadających odpowiednio: cechę X na poziomie i oraz cechę Y na poziomie j; n – liczba wszystkich wyników. Zakładając, że wartości cechy X zostały podzielone na r grup a wartości cechy Y na c grup liczebności te wyznacza się z wzorów:

$$n_i = \sum_{j=1}^{c} n_{ij} , \quad n_j = \sum_{i=1}^{r} n_{ij} , \quad n = \sum_{i=1}^{r} n_i = \sum_{j=1}^{c} n_j .$$

Dla wygody macierz liczebności empirycznych uzupełnia się o dodatkowe kolumny i wiersze w których wpisywane są liczebności n_i, n_j oraz prawdopodobieństwa p_i, p_j.

10+10 = 20

Trwałość w dniach	Metoda 1	Metoda 2	n_i	p_i
40-50	10	10	20	0.1
50-60	30	110	140	0.7
60-70	20	20	40	0.2
n_j	60	140	200	
p_j	0.3	0.7		1.0

20/200 = 0.1

n = 20+140+40 = 60+140

10+30+20 = 60

0.3+0.7=0.1+0.7+0.2=1.0

Następnie wyznaczane są prawdopodobieństwa teoretyczne p_{ij} i w oparciu o nie – liczebności teoretyczne: np_{ij}. Prawdopodobieństwa p_{ij} wpisywane są w prawy górny narożnik każdej kratki tablicy, liczebności np_{ij} – w lewy dolny narożnik.

0.1 0.3= 0.03

Trwałość w dniach	Metoda 1	Metoda 2	n_i	p_i
40-50	0.03 10 6	0.07 10 14	20	0.1
50-60	0.21 30 42	0.49 110 98	140	0.7
60-70	0.06 20 12	0.14 20 28	40	0.2
n_j	60	140	200	
p_j	0.3	0.7		1.0

Tak wypełniona tablica jest nazywana *tablicą niezależności*. Tablica ta jest też tworzona w wersji skróconej – bez obliczania prawdopodobieństw: p_i, p_j i p_{ij}. W tym przypadku wykorzystywany jest fakt, że liczebności teoretyczne można wyznaczyć z zależności:

$$n\,p_{ij} = n\,p_i\,p_j = n\frac{n_i}{n}\frac{n_j}{n} = \frac{n_i\,n_j}{n} \quad .$$

Tablica niezależności stanowi podstawę do obliczenia wartości statystyki testowej χ^2:

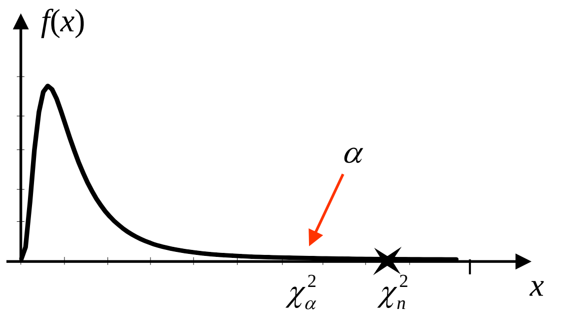
$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij} - np_{ij}\right)^2}{np_{ij}} .$$

Obliczenia te przeprowadzane są zwykle w formie tabelarycznej:

	n_{ij}	np_{ij}	$\left(n_{ij} - np_{ij}\right)^2$	$\left(n_{ij} - np_{ij}\right)^2 / np_{ij}$
	10	6	16	2.67
	10	14	16	1.14
	30	42	144	3.43
	110	98	144	1.47
	20	12	64	5.33
	20	28	64	2.29
Σ	**200**	**200**		**16.33**

Obliczona na podstawie wyników z próby wartość statystyki testowej wyniosła więc: $\chi_n^2 \approx 16.33$.

Obszar krytyczny wyznacza się wykorzystując rozkład χ^2 o $(3-1)(2-1) = 2$ stopniach swobody:



$$\chi_\alpha^2 = F_{\chi^2(2)}^{-1}(1-\alpha) = F_{\chi^2(2)}^{-1}(0.99) \approx 9.21$$

Wartość statystyki testowej leży w obszarze krytycznym, hipotezę H_0 należy więc odrzucić na rzecz hipotezy alternatywnej – trwałość produktu jest więc zależna od użytej metody produkcji.

Identyczny wniosek można wyciągnąć wyznaczając *graniczny poziom istotności p–value*, który w tym przypadku wynosi (test jest prawostronny):

$$p\text{–}value = 1 - F_{\chi^2(2)}\left(\chi_n^2\right) = 1 - F_{\chi^2(2)}(16.33) \approx 0.000284\,.$$