

SdI30 LABORATORIUM 11

Zestaw zadań W10 Testy nieparametryczne

Niezbędne tablice statystyczne

1. Tablica z wartościami dystrybucyjnymi rozkładu $\mathcal{N}(0; 1)$.
2. Tablica z kwantylami rozkładu normalnego.
3. Tablica z kwantylami rozkładu chi-kwadrat.

Zadania

1. Firma rozważa pięć projektów nazw swojego nowego produktu. Przed wybraniem jednej z nich firma postanowiła sprawdzić, czy wszystkie pięć nazw równie silnie przyciąga klientów. Wybrano próbę losową 100 osób i każdą z nich poproszono o wskazanie najlepszej spośród pięciu nazw. Liczby osób, które wybrały kolejne nazwy są podane poniżej:

Nazwa	A	B	C	D	E
a)	8	16	30	34	12
b)	4	12	34	40	10
c)	12	30	15	15	28

Przeprowadź test.

Odp.: b) $\chi_0^2 = 50,8$, odrzucamy hipotezę zerową.

2. Istnieje przekonanie, że zwroty z pewnej inwestycji mają rozkład normalny z wartością oczekiwaną 11% (roczna stopa) i odchyleniem standardowym 2%. Firma brokerska chcąc przetestować hipotezę zerową, że przekonanie to jest prawdziwe, zebrała następujące dane o zwrotach w % (zakładamy, że jest to próba losowa):
 - i) 8,0; 9,0; 9,5; 9,5; 8,6; 13,0; 14,5; 12,0; 12,4; 19,0; 9,0; 10,0; 10,0; 11,7; 15,0; 10,1; 12,7; 17,0; 8,0; 9,9; 11,0; 12,5; 12,8; 10,6; 8,8; 9,4; 10,0; 12,3; 12,9; 7,0.
 - ii) Do wszystkich wartości dodajemy 0,2.
 - iii) Do nieparzystych wartości dodajemy 0,5, a od parzystych wartości odejmujemy 0,5.
 - a) Przeprowadź analizę statystyczną i sformułuj wniosek.
 - b) Przeprowadź test hipotezy zerowej, że zwroty z inwestycji mają rozkład normalny, ale z nieznaną wartością oczekiwaną.
 - c) Przeprowadź test hipotezy zerowej, że zwroty z inwestycji mają rozkład normalny, ale z nieznaną wartością oczekiwaną i nieznanym odchyleniem standardowym.

Odp.: i) c) $\bar{x} = 11,21$, $s = 2,71$, statystyka chi-kwadrat jest na tyle mała, że nie możemy odrzucić hipotezy zerowej.

3. Zaplanować doświadczenie do zbadania symetryczności monety. Powiedzmy, że w 100 rzutach monetą otrzymano

- a) 55 orłów,
- b) 65 orłów.

Co możemy powiedzieć o symetryczności monety na poziomie istotności 0,05?

Odp.: a) $\chi_0^2 = 1$; **b)** $\chi_0^2 = 9$; $\chi_{0,95;1}^2 = 3,8415$.

4. Zaplanować doświadczenie do zbadania prawidłowości kostki do gry. Powiedzmy, że w 120 rzutach kostką otrzymano następujące wyniki:

Liczba oczek	1	2	3	4	5	6
i)	14	18	23	22	28	15
Liczba rzutów ii)	11	30	14	10	33	22

Co możemy powiedzieć o prawidłowości kostki na poziomie istotności 0,05?

Odp.: ii) $\chi_0^2 = 24,5 > \chi_{0,95;5}^2 = 11,0705$.

5. Do każdej z 20 tarcz oddano po 5 niezależnych strzałów i zanotowano liczbę trafień. Wyniki strzelania podane są w tabeli:

Liczba trafień	0	1	2	3	4	5
i)	0	2	8	6	3	1
Liczba tarcz ii)	1	2	3	10	3	1

Na poziomie istotności 0,1 zweryfikować hipotezę orzekającą, że wyniki strzelania mają rozkład dwumianowy.

Odp.: ii) Łączymy klasy i szacujemy parametr p , $\hat{p} = 0,55$. Odrzucamy hipotezę.

6. W celu zbadania rozkładu liczby awarii w sieci wodno-kanalizacyjnej pewnego miasta przeprowadzono obserwacje w ciągu 100 dni. Wyniki obserwacji podane są w tabeli.

Dzienna liczba awarii	0	1	2	3	4
i)	22	30	22	16	10
Liczba dni ii)	10	22	30	22	16

Na poziomie istotności 0,05 zbadać, czy rozkład liczby awarii jest rozkładem równomiernym.

7. Dane z próby zostały pogrupowane w tabeli:

Przedział	Liczba wyników	
	i)	ii)
$(-\infty, -1)$	11	31
$[-1, 1]$	76	96
$(1, \infty)$	13	33

Na poziomie istotności 0,05 zweryfikować hipotezę, że dane pochodzą ze standardowego rozkładu normalnego.

Odp.: i) $\chi_0^2 = 2,89 < \chi_{0,95;2}^2 = 5,9915$.

8. Przeprowadzono badanie wytrzymałości betonu na ściskanie. Uzyskane wyniki pomiarów (w N/cm^2) są podane w tabeli:

Wytrzymałość	Liczba próbek	
	i)	ii)
(1900 – 2000]	14	10
(2000 – 2100]	26	26
(2100 – 2200]	52	56
(2200 – 2300]	58	64
(2300 – 2400]	33	30
(2400 – 2500]	17	14

Na poziomie istotności 0,05 sprawdzić, czy wytrzymałość betonu na ściskanie

- ma rozkład normalny;
- ma rozkład $\mathcal{N}(2200; \sigma)$;
- ma rozkład $\mathcal{N}(2200; 100)$.

Odp.: ii) a) Szacujemy dwa parametry, stąd 3 stopnie swobody, $\chi_0^2 = 2,91 < \chi_{0,95;3}^2 = 7,8147$.

9. Dane z próby zostały pogrupowane w tabeli:

Przedział	(0, 1]	(1, 2]	(2, 3]	(3, 4]	(4, 5]	(5, 6]	(6, 7]	(7, 8]	(8, 9]	(9, 10]
liczba i)	52	38	20	12	7	6	5	5	4	1
wyników ii)	33	36	19	14	9	8	4	6	5	4

Na poziomie istotności 0,02 zweryfikować hipotezę, że dane te pochodzą z rozkładu o gęstości $f(x)$ określonej wzorem (wyznaczyć a):

$$f(x) = a(10 - x)\mathbb{I}_{[0;10]}(x)$$

10. Badaniu poddano ciało radioaktywne ze względu na ilość emitowanych przez nie cząstek. Badanie polegało na obserwacji tego ciała w ciągu 2608 jednakowych okresów (po 7,5 sekundy każdy). Dla każdego okresu rejestrowano ilość cząstek wpadających do licznika. Wynik rejestracji jest zestawiony w postaci ciągu par i, n_i , gdzie $i = 0, 1, 2, \dots, 10$ oznacza liczbę cząstek wpadających do licznika, natomiast n_i liczbę okresów, w których zaobserwowano i cząstek wpadających do licznika:

[0, 57], [1, 203], [2, 383], [3, 525], [4, 532], [5, 408], [6, 273], [7, 139], [8, 45], [9, 27], [10, 16].

Zbadać, czy ilość cząstek emitowanych przez badane ciało radioaktywne jest zgodna z rozkładem Poissona. Odp.: Chisquare = 20.6105 with 10 d.f. Sig. level = 0.023979

11. (KA 6.18). Wynikami pięcioelementowej próby są: 1.37, 0.18, 0.56, 2.46, 0.87. Zapisać wyniki w zmiennej *wyn*. Na poziomie istotności $\alpha = 0,05$ zweryfikować hipotezę, że próba została pobrana z populacji $X \sim \exp(1)$.

12. (KA 6.19). Sprawdzić na poziomie istotności $\alpha = 0,05$ hipotezę, że próby *pba1* i *pba2*, gdzie

$pba1 = \{0.46, 0.14, 2.45, -0.32, -0.07, 0.3\}$,

$pba2 = \{0.06, -2.53, -0.53, -0.19, 0.54, -1.56, 0.19, -1.19, 0.02\}$

pochodzą z populacji X i Y o tym samym rozkładzie.

Odp.: zaobserwowana wartość statystyki $KS2(x; y) = 8/18$

13. Wygenerować próby o liczebności 100 obserwacji według rozkładów:

i) $\mathcal{N}(900; 50)$,

ii) $TR(725; 1075)$,

Następnie

- obliczyć podstawowe statystyki,
- sporządzić wykresy histfit, normplot, Q-Q,
- przeprowadzić testy losowości,
- przeprowadzić testy normalności,
- przeprowadzić testy zgodności z innymi rozkładami,
- przeprowadzić test zgodności dla wygenerowanych prób.

14. Populacja pewnych elementów badana jest ze względu na dwie cechy, cechę Y przyjmującą wartości z przedziału (0, 1) oraz cechę X przyjmującą tylko dwie wartości 0 i 1. Dane z próby losowej zebrane w czteropolowej tablicy są następujące:

$X \setminus Y$	$< 0,5$	$\geq 0,5$
0	72	29
1	53	26

Na poziomie istotności 0,05 zweryfikować hipotezę o niezależności cech X i Y .

Odp.: Nie ma podstaw do odrzucenia hipotezy o niezależności badanych cech.

15. Dla zbadania wpływu palenia (zmienna X) na zachorowania na raka (zmienna Y) wylosowano 874 osoby z dorosłej populacji ludzi i uzyskano następujące dane:

$X \setminus Y$	ma raka	nie ma raka
pali	412	299
nie pali	32	131

Na poziomie istotności 0,02 zweryfikować hipotezę o niezależności palenia i zachorowalności na raka.

Odp.: Odrzucamy hipotezę o niezależności palenia i zachorowalności na raka.

16. Wyrób produkowany w dwóch zakładach A i B może być uznany jako wadliwy z dwóch przyczyn: J – niskiej jakości wykonania lub S – użycia gorszego surowca. Analizując losową próbę wyrobów uzyskano wyniki podane w tablicy:

$X \setminus Y$	A	B
J	33	46
S	16	5

Na poziomie istotności 0,01 zweryfikować hipotezę o niezależności między miejscem powstania wyrobu (zm. Y) a przyczyną uznania wyrobu za wadliwy (zm. X).

17. Zbadać, czy udziały w rynku firm A , B , C , D , E wynajmujących samochody zmieniły się, jeśli dane z dwóch lat dla prób losowych są następujące:

Rok \ Firma	A	B	C	D	E
I	39	26	18	14	3
II	29	25	16	19	11

18. Wygenerować dużą próbę według jednego z rozkładów: beta, gamma, Weibulla lub logarytmiczno-normalnego i przekazać uzyskane dane drugiej osobie do identyfikacji rozkładu – nie informując o mechanizmie generowania. Dokonać oceny jakości dokonanej identyfikacji.