

ANALIZA DANYCH ANKIETOWYCH

Lista 5, 6 i 7

Zadania do sprawozdania 2, część 1

1. Zapoznać się z funkcjami *binom.test* i *prop.test*.
2. Załóżmy, że 200 losowo wybranych klientów (w różnym wieku) kilku (losowo wybranych) aptek zapytano, jaki lek przeciwbólowy zwykle stosują. Zebrane dane zawarte są w tablicy 1 (są to te same dane, które były rozważane w zadaniach z listy 3 i 4).

Tablica 1: Dane do zadania 2. i 3.

Lek	Wiek ankietowanych			Suma
	do lat 35	od 36 do 55	powyżej 55	
Ibuprom	35	0	0	35
Apap	22	22	0	44
Paracetamol	15	15	15	45
Ibuprofen	0	40	10	50
Panadol	18	3	5	26
Suma	90	80	30	200

Na podstawie tych danych, na poziomie istotności $\alpha = 0.05$, zweryfikować następujące hipotezy:

- (a) prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Apap jest mniejsze bądź równe $1/4$,
- (b) prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Apap jest równe $1/2$,
- (c) prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Ibuprom jest większe bądź równe $1/5$,
- (d) powtórzyć punkt (a), (b) i (c), ale dla osoby z badanej populacji do lat 35.

W każdym z powyższych punktów podać wartość poziomu krytycznego (p value) i sformułować odpowiedź.

3. Na podstawie danych w tablicy 1, korzystając z testu Fishera, na poziomie istotności $\alpha = 0.05$, zweryfikować hipotezę, że prawdopodobieństwo, że osoba do lat 35 zażywa Panadol jest równe prawdopodobieństwu, że osoba od 36 lat do 55 lat zażywa Panadol.

Czy na podstawie uzyskanego wyniku można (na zadanym poziomie istotności) odrzucić hipotezę o niezależności wyboru leku Panadol w leczeniu bólu od wieku, przy uwzględnieniu tylko dwóch grup wiekowych - do lat 35 i od 36 do 55 lat?

4. Korzystając z funkcji *chisq.test* w pakiecie R, na poziomie istotności 0.05, zweryfikować hipotezę o niezależności stopnia zadowolenia z pracy i wynagrodzenia na podstawie danych w tablicy 2. Zwrócić uwagę na stosowaną w tej funkcji poprawkę.

Tablica 2: Dane do zadania 4.

Wynagrodzenie	Stopień zadowolenia z pracy				Suma
	b. niezadow.	niezadow.	zadow.	b. zadow.	
do 6000	20	24	80	82	206
6000–15000	22	38	104	125	289
15000-25000	13	28	81	113	235
powyżej 25000	7	18	54	92	171
Suma	62	108	319	412	901

5. Napisać deklarację funkcji, która dla danych w tablicy dwudzielczej oblicza wartość poziomu krytycznego (p-value) w asymptotycznym teście niezależności opartym na ilorazie wiarygodności. Korzystając z napisanej funkcji, obliczyć tę wartość dla danych z zadania 4.

Zadanie dodatkowe

Test niezależności (ogólny, nie tylko dla danych wielomianowych) oparty na korelacji odległości jest częścią pakietu *energy* w R. Jednak, według mnie, zwraca niewłaściwe wyniki. Napisać funkcję, która dla dwóch wektorów danych (danych surowych, a nie zebranych w tabeli dwudzielczej) oblicza wartość poziomu krytycznego (p-value) w teście opartym na korelacji odległości. Następnie dla wygenerowanych danych (dwóch wektorów - realizacji niezależnych danych dyskretnych) zweryfikować hipotezę o niezależności przy użyciu napisanej funkcji i korzystając z funkcji dostępnej w pakiecie.

Alicja Jokiel-Rokita

24 marca 2022