

ANALIZA DANYCH ANKIETOWYCH

Zadania do sprawozdania 2

Testowanie niezależności

1. W roku 1986 miała miejsce katastrofa promu kosmicznego Challenger, której przyczyną było uszkodzenie pierścieni uszczelniających. Start promu odbywał się w temperaturze 62°F . W celu odpowiedzi na pytanie, czy można było uniknąć tej tragedii, przeanalizujemy dane z 24 wcześniejszych startów wahadłowców w przestrzeń kosmiczną. Każdy start można opisać dwiema cechami: temperaturą otoczenia zdyskretyzowaną do dwóch przedziałów (co najwyżej 65°F i powyżej 65°F) oraz uszkodzeniami pierścieni uszczelniających (brak uszkodzeń i obecność co najmniej jednego uszkodzenia). Dane są następujące: wszystkim czterem startom w temperaturze nie przekraczającej 65°F towarzyszyły uszkodzenia pierścieni, natomiast w 17 z 20 startów w wyższych temperaturach nie nastąpiło uszkodzenie żadnego z pierścieni. Czy na poziomie istotności 0.05 są podstawy do odrzucenia hipotezy o niezależności temperatury i wystąpienia uszkodzenia? Uzasadnić przyjętą metodę wnioskowania.
 2. Dane w pliku *Reakcja.csv* zawierają informację o reakcji na lek (zmienna *Reakcja* na poziomie 0, gdy nie nastąpiła poprawa i na poziomie 1, gdy nastąpiła poprawa) w różnych dawkach (zmienna *Dawka* (w skali logarytmicznej)), dwóch firm farmaceutycznych (zmienna *Rodzaj*) pacjentów leczonych w domu (0) lub w szpitalu (1) (zmienna *Miejsce*). Na podstawie uzyskanych danych odpowiedzieć na następujące pytania.
 - (a) Czy skuteczność leczenia jest niezależna od wielkości dawki?
 - (b) Czy skuteczność leczenia jest niezależna od rodzaju leku?
 - (c) Czy skuteczność leczenia jest niezależna od miejsca leczenia?Opisać i uzasadnić przyjętą metodę wnioskowania.
 3. 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 1. Zwrócić uwagę na stosowaną w tej funkcji poprawkę.
 4. Napisać deklarację funkcji, która dla danych w tablicy dwudzielczej oblicza wartość poziomu krytycznego (p-value) w teście niezależności opartym na ilorazie wiarygodności. Korzystając z napisanej funkcji, obliczyć tę wartość dla danych z zadania 3.
-

Miary współzmienności i analiza korespondencji

5. W oparciu o dane w pliku *Reakcja.csv* obliczyć wartości miar współzmienności zmiennych *Reakcja* (skuteczność leczenia) i *Dawka* (wielkość dawki) oraz *Reakcja* i *Miejsce* (miejsce leczenia). Podać interpretację tych wartości.

Tablica 1: Dane do zadania 3.i 6.

Wynagrodzenie	Stopień zadowolenia z pracy				Suma
	b. niezadow.	niezadow.	zadow.	b. zadow.	
poniżej 6000	32	44	60	70	206
6000-15000	22	38	104	125	289
15000-25000	13	48	61	113	235
powyżej 25000	3	18	54	96	171
Suma	62	108	319	412	901

6. Na podstawie danych zawartych w tablicy 1, obliczyć (odpowiednią) miarę współzmienności zmiennych *Wynagrodzenie* i *Stopień zadowolenia z pracy*. Następnie, przeprowadzić analizę korespondencji, tzn. obliczyć wartości odpowiednich macierzy, współrzędnych punktów oraz utworzyć odpowiednie wykresy.
7. Załóżmy, że 200 klientów (w różnym wieku) kilku aptek zapytano, jaki lek przeciwbólowy zwykle stosują. Zebrane dane zawarte są w tablicy 2. Na podstawie tych danych, obliczyć odpowiednie miary współzmienności oraz przeprowadzić analizę korespondencji, tzn. obliczyć wartości odpowiednich macierzy, współrzędnych punktów oraz utworzyć odpowiednie wykresy.

Tablica 2: Dane do zadania 7.

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

Alicja Jokiel-Rokita

3 listopada 2023