
ANALIZA DANYCH ANKIETOWYCH, SEMESTR LETNI 2023/2024

Zadania do sprawozdania 2

Część I

zadanie 1. W ankiecie przedstawionej na poprzedniej liście pracownicy zostali poproszeni o wyrażenie opinii na temat podejścia firmy do utrzymania równowagi między życiem zawodowym a prywatnym. Wśród próbki 200 pracowników (losowanie proste ze zwracaniem) uzyskano wyniki:

- 14 pracowników - bardzo niezadowolonych,
- 17 pracowników - niezadowolonych,
- 40 pracowników - nie ma zdania,
- 100 pracowników - zadowolonych,
- 29 pracowników - bardzo zadowolonych,

Na podstawie danych wyznacz przedział ufności dla wektora prawdopodobieństw opisującego stopień zadowolenia z podejścia firmy. Przyjmij poziom ufności 0.95.

zadanie 2. Napisz funkcję, która wyznacza wartość poziomu krytycznego w następujących testach:

- chi-kwadrat Pearsona
- chi-kwadrat największej wiarygodności

służących do weryfikacji hipotezy $H_0 : \mathbf{p} = \mathbf{p}_0$ przy hipotezie alternatywnej $H_0 : \mathbf{p} \neq \mathbf{p}_0$ na podstawie obserwacji \mathbf{x} wektora losowego \mathbf{X} z rozkładu wielomianowego z parametrami n i \mathbf{p} .

zadanie 3. Na podstawie danych z ankiety z poprzedniej listy zweryfikuj hipotezę, że w grupie pracowników zatrudnionych w Dziale Kreatywnym rozkład odpowiedzi na pytanie dotyczące podejścia firmy do utrzymania równowagi między życiem zawodowym a prywatnym jest równomierny, tzn. jest jednakowe prawdopodobieństwo, że pracownik zatrudniony w Dziale Kreatywnym jest udzielił odpowiedzi "zdecydowanie się nie zgadzam", "nie zgadzam się", "nie mam zdania", "zgadzam się", "zdecydowanie się zgadzam" na pytanie PYT_1. Przyjmij poziom istotności 0.05. Skorzystaj z funkcji napisanej w zadaniu 2.

Część II

zadanie 4. Zapoznaj się z funkcją *fisher.test* z pakietu *stats*.

zadanie 5. Korzystając z testu Fishera, na poziomie istotności 0.05, zweryfikuj hipotezę, że zmienna **PŁEĆ** i zmienna **CZY_KIER** są niezależne. Czy na poziomie istotności 0.05 możemy wnioskować, że prawdopodobieństwo tego, że na stanowisku kierowniczym pracuje kobieta jest równe prawdopodobieństwu tego, że na stanowisku kierowniczym pracuje mężczyzna?

zadanie 6. Korzystając z testu Freemana-Haltona na poziomie istotności 0.05 zweryfikuj następujące hipotezy:

- zajmowanie stanowiska kierowniczego nie zależy od wieku (**CZY_KIER** oraz **WIEK_KAT**),
- zajmowanie stanowiska kierowniczego nie zależy od stażu pracy (**CZY_KIER** oraz **STAŻ**),
- zadowolenie z wynagrodzenia w pierwszym badanym okresie nie zależy od zajmowanego stanowiska (**PYT_2** oraz **CZY_KIER**),
- zadowolenie z wynagrodzenia w pierwszym badanym okresie nie zależy od stażu (**PYT_2** oraz **STAŻ**),
- zadowolenie z wynagrodzenia w pierwszym badanym okresie nie zależy od płci (**PYT_2** oraz **PŁEĆ**),
- zadowolenie z wynagrodzenia w pierwszym badanym okresie nie zależy od wieku (**PYT_2** oraz **WIEK_KAT**),

Część III

zadanie 7. Zapoznaj się z funkcją *chisq.test* z pakietu *stats* oraz *assocplot* z pakietu *graphics*.

zadanie 8. Korzystając z *chisq.test* zweryfikuj hipotezę, że zadowolenie z wynagrodzenia w pierwszym badanym okresie nie zależy od zajmowanego stanowiska. Przyjmij poziom istotności 0.01. Stwórz wykres przy pomocy funkcji *assocplot* i dokonaj jego interpretacji. Wynik testu porównaj z wynikiem uzyskanym w zadaniu 6.

zadanie 9. Zapoznaj się z funkcją *rmultinom* z pakietu *stats*, a następnie korzystając z niej przeprowadź symulacje w celu oszacowania mocy testu Fishera oraz mocy testu chi-kwadrat Pearsona, generując dane z tabeli 2×2 , w której $p_{11} = 1/40$, $p_{12} = 3/40$, $p_{21} = 19/40$, $p_{22} = 17/40$. Symulacje wykonaj dla $n = 50$, $n = 100$ oraz $n = 1000$.

zadanie 10. Napisz funkcję, która dla danych z tablicy dwudzielczej oblicza wartość poziomu krytycznego w teście niezależności opartym na ilorazie wiarygodności. Korzystając z napisanej funkcji, wykonaj test dla danych z zadania 8.

Część IV oraz V

zadanie 11. Przeprowadzone wśród brytyjskich mężczyzn badanie trwające 20 lat wykazało, że odsetek zmarłych (na rok) z powodu raka płuc wynosił 0,00140 wśród osób palących papierosy i 0,00010 wśród osób niepalących. Odsetek zmarłych z powodu choroby niedokrwiennej serca wynosił 0,00669 dla palaczy i 0,00413 dla osób niepalących. Opisz związek pomiędzy paleniem papierosów a śmiercią z powodu raka płuc oraz związek pomiędzy paleniem papierosów a śmiercią z powodu choroby serca. Skorzystaj z różnicy proporcji, ryzyka względnego i ilorazu szans. Zinterpretuj wartości. Związek której pary zmiennych jest silniejszy?

zadanie 12. Tabela 1 przedstawia wyniki dotyczące śmiertelności kierowców i pasażerów w wypadkach samochodowych na Florydzie w 2008 roku, w zależności od tego, czy osoba miała zapięty pas bezpieczeństwa czy nie.

Tabela 1: Wyniki dotyczące śmiertelności w wypadkach samochodowych na Florydzie w 2008 roku

	Śmiertelny	Nieśmiertelny
Bez pasów	1085	55,623
Z pasami	703	441,239

- Oszacuj warunkowe prawdopodobieństwo śmierci w wypadku ze względu na drugą zmienną, tj. dla kierowców i pasażerów, którzy użyli pasa bezpieczeństwa oraz dla kierowców i pasażerów, którzy nie użyli pasa bezpieczeństwa.
- Oszacuj warunkowe prawdopodobieństwo użycia pasa bezpieczeństwa ze względu na drugą zmienną, tj. dla kierowców i pasażerów ze śmiertelnymi obrażeniami oraz dla kierowców i pasażerów, którzy przeżyli wypadek.
- Jaki jest najbardziej naturalny wybór dla zmiennej objaśnianej w tym badaniu? Dla takiego wyboru wyznacz i zinterpretuj różnicę proporcji, ryzyko względne oraz iloraz szans. Dlaczego wartości ryzyka względnego i ilorazu szans przyjmują zbliżone wartości?

zadanie 13. Oblicz wartości odpowiednich miar współzmienności (współczynnik tau lub współczynnik gamma) dla zmiennych:

- zadowolenie z wynagrodzenia w pierwszym badanym okresie i zajmowane stanowisko,
- zadowolenie z wynagrodzenia w pierwszym badanym okresie i staż pracy,
- zajmowane stanowisko i staż pracy.

zadanie 14. Na podstawie informacji przedstawionych na wykładzie napisz własną funkcję do przeprowadzania analizy korespondencji. Funkcja powinna przyjmować jako argument tablicę dwudzielczą i zwracać obliczone wartości odpowiednich wektorów i macierzy, współrzędnych punktów oraz odpowiedni wykres. Korzystając z napisanej funkcji wykonaj analizę korespondencji dla danych przedstawionych w na wykładzie.

Zadania dodatkowe

zadanie *1. Napisz funkcję, która dla dwóch wektorów danych oblicza wartość poziomu krytycznego (p-value) w teście opartym na korelacji odległości. Następnie dla wygenerowanych danych zweryfikuj hipotezę o niezależności przy użyciu napisanej funkcji.