Raport 2

Natalia Iwańska 262270, Klaudia Janicka 262268

2023-05-12

Zadanie 2

Przy pomocy testu Fishera na poziomie istotności $\alpha = 0.05$ zweryfikowano następującą hipotezę:

$H_0:$ płeć i zajmowane stanowisko nie zależą od siebie,

przeciwko

$H_1:$ płeć i zajmowane stanowisko są od siebie zależne.

	K	M
nie	63.00	110.00
tak	8.00	19.00

Tab. 1: Tablica dwudzielcza.

Wnioski

Na zadanym poziomie istotności, $\alpha = 0.05$, wyliczona p-wartość wynosi 0.6659029 co sugeruje, że nie ma podstaw do odrzucenia hipotezy zerowej. Zatem można przyjąć, że prawdopodobieństwo, że na stanowisku kierowniczym pracuje kobieta jest równe prawdopodobieństwu, że na stanowisku kierowniczym pracuje mężczyzna.

Zadanie 3

Korzystając z testu Freemana-Haltona, na poziomie istotności $\alpha = 0.05$, zweryfikowano następujace hipotezy:

- $H_0:$ Zajmowanie stanowiska kierowniczego nie zależy od wieku.

	0	1
1	23	3
2	91	13
3	39	6
4	20	5

Tab. 2: Tablica dwudzielcza dla stanowiska oraz wieku.

W przeprowadzonym teście p-wartość wyniosła 0.7823002.

- $H_0:$ Zajmowanie stanowiska kierowniczego nie zależy od wykształcenia.

	0	1
1	40	1
2	123	17
3	10	9

Tab. 3: Tablica dwudzielcza dla stanowiska oraz wykształcenia.

W przeprowadzonym teście p-wartość wyniosła 6.5378957×10^{-5}.

Wnioski

w pierwszym przeprowadzonym teście wyliczona p-wartość sugeruje, że nie ma podstaw do odrzucenia hipotezy zerowej, iż zajmowanie stanowiska kierowniczego nie zależy od wieku. Natomiast w 2. przypadku odrzucono hipotezę zerową - zajmowanie stanowiska kierowniczego nie zależy od wykształcenia.

zadanie 4

Korzystając z testu Freemana-Haltona, na poziomie istotności $\alpha = 0.05$, zweryfikowano następujace hipotezy:

- H_0: Zadowolenie z wynagrodzenia (w pierwszym badanym okresie) nie zależy od zajmowanego stanowiska.

	-2	-1	1	2
0	64	18	0	91
1	10	2	2	13

Tab. 4: Tablica dwudzielcza dla zadowolenia z wynagrodzenia oraz stanowiska.

Wnioski

W przeprowadzonym teście p-wartość wyniosła 0.0442973. Zatem hipotezę, że zadowolenie z wynagrodzenia nie zależy od zajmowanego stanowiska należy odrzucić.

- H_0: Zadowolenie z wynagrodzenia (w pierwszym badanym okresie) nie zależy od wykształcenia.

	-2	-1	1	2
1	20	3	0	18
2	45	17	0	78
3	9	0	2	8

Tab. 5: Tablica dwudzielcza dla zadowolenia z wynagrodzenia oraz wykształcenia.

Wnioski

W przeprowadzonym teście p-wartość wyniosła 0.0106902. Na tej podstawie odrzucono hipotezę o niezależności zadowolenia z wynagrodzenia od wykształcenia.

- H_0: Zadowolenie z wynagrodzenia(w pierwszym badanym okresie) nie zależy od płci.

	-2	-1	1	2
K	25	10	1	35
M	49	10	1	69

Tab. 6: Tablica dwudzielcza dla zadowolenia z wynagrodzenia oraz płci.

Wnioski

Na podstwie otrzymanej p-wartości (0.4758086) nie ma podstw do odrzucenia hipotezy o niezależności zadowolenia z wynagrodzenia od płci.

- H_0: Zadowolenie z wynagrodzenia (w pierwszym badanym okresie) nie zależy od wieku.

	-2	-1	1	2
1	9	1	0	16
2	42	9	1	52
3	12	6	0	27
4	11	4	1	9

Tab. 7: Tablica dwudzielcza dla zadowolenia z wynagrodzenia oraz wieku.

Wnioski

Bazując na p-wartość, która wyniosła 0.319352 nie ma podstaw do odrzucenia hipotezy o niezależności zadowolenia z wynagrodzenia od wieku.

Zadanie 6

Korzystając z testu chi-kwadrat Pearsona i z testu chi-kwadrat ilorazu wiarogodności na poziomie istotności $\alpha = 0.01$ zweryfikowano następującą hipotezę

H_0: Zadowolenie z wynagrodzenia nie zależy od zajmowanego stanowiska.

		0	1
	1	64	10
g	2	18	2
	3	0	2
	4	91	13

Tab. 8: Tablica dwudzielcza dla zadowolenia z wynagrodzenia oraz zajmowanego stanowiska.

Tab. 9: P-wartości dla poszczególnych testów.

test	p.wartość
chi-kwadrat Pearsona	0.0043971
chi-kwadratilorazu wiarogodności	0.0396896

Wnioski

P-wartość przeprowadzonego testu chi-kwadrat Pearsona wynosi 0.0043971, co oznacza, że hipotezę zerową na poziomie ufności 0.01 należy odrzucić. Natomiast p-wartość testu chi-kwadratu ilorazu wiarygodności jest równa 0.03968965 co oznacza, że nie ma podstaw do odrzucenia hipotezy zerowej na rzecz alternatywnej.

W zadaniu 4a, jak i w tym, badamy te same hipotezy, ale na innych poziomach istotności. W zadaniu 4a, jak również i z wykorzystaniem testu chi-kwadrat ilorazu wiarygodności nie odrzucamy hipotezy zerowej. **(w 4a odrzucamy H_0 -> trzeba inny wniosek)**

Zadanie 7

W celu oszacowania zarówno rozmiaru jak i mocy testu odpowiednio dla testu Fishera, testu chi-kwadrat Pearsona oraz testu ilorazu wiarogodności przeprowadzone zostały symulacje Monte Carlo z liczbą powtórzeń $M = 5000$. Wszystkie testy zostały przeprowadzone na poziomie istotności $\alpha = 0.05$. Wymienione powyżej testy weryfikują hipotezę o niezależności

H_0: $\boldsymbol{p} \in \mathcal{P}_0$, gdzie $\mathcal{P}_0 = \{\boldsymbol{p} = (p_{11}, \ldots, p_{1C}, \ldots, p_{RC}) : p_{ij} = p_{i+}p_{+j}\}$.

a) Szacowanie rozmiaru testu

W poniższych symulacjach wektor prawdopodobieństw w rozkładzie wielomianowym jest postaci $p = \left(\frac{1}{20}, \frac{9}{20}, \frac{1}{20}, \frac{9}{20}\right)$. Wektor ten można zapisać w postaci tablicy 2x2 wraz z prawdopodobieństwami brzegowymi.

	1	2	i+
1	0.05	0.45	0.5
2	0.05	0.45	0.5
+j	0.10	0.90	1.0

W celu weryfikacji zgodności powyższego wektora z hipotezą zerową sprawdzono, czy warunek $p_{ij} = p_{i+}p_{+j}$ jest spełniony dla każdego i, j z zadanego wektora.

- $p_{11} = 0.05 \quad = \quad p_{1+}p_{+1} = 0.5 \cdot 0.10 = 0.05$
- $p_{12} = 0.45 \quad = \quad p_{1+}p_{+2} = 0.5 \cdot 0.90 = 0.45$
- $p_{21} = 0.05 \quad = \quad p_{2+}p_{+1} = 0.5 \cdot 0.10 = 0.05$
- $p_{22} = 0.45 \quad = \quad p_{2+}p_{+2} = 0.5 \cdot 0.90 = 0.45$

Zatem podany wektor prawdopodobieństw jest zgodny z hipotezą zerową.

Tab. 10: Prawdopodobieństwo popełnienie błędu I rodzaju w zależności od rozmiaru próby.

n	test Fishera	test chi-kwadrat Pearsona	test ilorazu wiarogodności
50	0.0074	0.0032	0.0218
100	0.0238	0.0130	0.0478
1000	0.0476	0.0422	0.0544

Wnioski

Prawdopodobieństwo popełnienia błędu I rodzaju zbliżone do $\alpha = 0.05$ dla testu Fishera oraz testu chi-kwadrat Pearsona otrzymano dopiero dla próby rozmiaru 1000. Natomiast w przypadku testu ilorazu wiarogodności zbliżony wynik osiągnięto już w przypadku $n = 100$, a dla $n = 1000$ poziom 0.05 został nieznacznie przekroczony. Ostatecznie można stwierdzić, że test Fishera oraz test chi-kwadrat Pearsona są testami na poziomie istotności α. **(?, plus nie wiem co z testem ilorazu wiarogodności)**

b) Szacowanie mocy testu

W poniższych symulacjach wektor prawdopodobieństw w rozkładzie wielomianowym jest postaci $p = \left(\frac{1}{40}, \frac{19}{40}, \frac{3}{40}, \frac{17}{40}\right)$. Ponownie wektor ten można zapisać w postaci tablicy 2x2 wraz z prawdopodobieństwami brzegowymi.

	1	2	i+
1	0.025	0.475	0.5
2	0.075	0.425	0.5
+j	0.100	0.900	1.0

W celu weryfikacji zgodności powyższego wektora z hipotezą zerową sprawdzono, czy warunek $p_{ij} = p_{i+}p_{+j}$ jest spełniony dla każdego i, j z zadanego wektora.

- $p_{11} = 0.025 \quad \neq \quad p_{1+}p_{+1} = 0.5 \cdot 0.10 = 0.05$

Zatem podany wektor prawdopodobieństw nie jest zgodny z hipotezą zerową.

Tab. 11: Moc testu w zależności od wielkości próby.

n	test Fishera	test chi-kwadrat Pearsona	test ilorazu wiarogodności
50	0.0432	0.0238	0.0946
100	0.2758	0.2120	0.3690
1000	0.9996	0.9994	0.9996

Wnioski

Największe prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest ona fałszywa dla każdej z testowanych długości próby odnotowano dla testu ilorazu wiarogodności. Zatem można wnioskować, że spośród badanych testów test ten ma największą moc i powinniśmy go stosować. **(? do sprawdzenia)**

Zadanie 8

Dla odpowiednich tabel dwudzielczych wyznaczono miary współzmienności.

- zadowolenie z wynagrodzenia (w pierwszym badanym okresie) i zajmowane stanowisko

Tab. 12: Tablica dwudzielcza dla zmniennej W1 oraz S.

	-2	-1	1	2	Sum
0	64	18	0	91	173
1	10	2	2	13	27
Sum	74	20	2	104	200

Dla powyższej tabeli przeprowadzono test Fishera na poziomie istotności $\alpha = 0.5$, w celu sprawdzenia hipotezy o niezależności. Na podstawie otrzymanej p-wartości (0.0443) odrzucono hipotezę o niezależności. Ponieważ zmienna S jest nominalna to do wyliczenia miary współzmienności wykorzystano współczynnik tau (współczynnik Goodmana i Kruskala), którego wartość wyniosła 0.0336229. Zatem można uznać, że **POTRZEBNE WNIOSKI**

- zadowolenie z wynagrodzenia (w pierwszym badanym okresie) i wykształcenie

Tab. 13: Tablica dwudzielcza dla zmniennej W1 oraz Wyk.

	-2	-1	1	2	Sum
1	20	3	0	18	41
2	45	17	0	78	140
3	9	0	2	8	19
Sum	74	20	2	104	200

Dla powyższej tabeli przeprowadzono test Fishera na poziomie istotności $\alpha = 0.5$, w celu sprawdzenia hipotezy o niezależności. Na podstawie otrzymanej p-wartości (0.0107) odrzucono hipotezę o niezależności. Ponieważ obie zmienne są porządkowe do wyliczenia miary współzmienności wykorzystano współczynnik gamma, którego wartość wyniosła 0.0908426. Zatem można uznać, że zmienne są ze sobą dodatnio zależne.

- zajmowane stanowisko i wykształcenie

Tab. 14: Tablica dwudzielcza dla zmniennej S oraz Wyk.

	0	1	Sum
1	40	1	41
2	123	17	140
3	10	9	19
Sum	173	27	200

Dla powyższej tabeli przeprowadzono test Fishera na poziomie istotności $\alpha = 0.5$, w celu sprawdzenia hipotezy o niezależności. Na podstawie otrzymanej p-wartości (10^{-4}) odrzucono hipotezę o niezależności. Ponieważ zmienna S jest nominalna to do wyliczenia miary współzmienności wykorzystano współczynnik tau (współczynnik Goodmana i Kruskala), którego wartość wyniosła 0.0732466. Zatem można uznać, że ... **POTRZEBNE WNIOSKI**

Zadanie 9

```
tabela <- as.matrix(ftable(personel$Wyk, personel$W1))

analiza.korespondencji <- function(tabela){
  n <- length(tabela[,1])
  m <- length(tabela[1,])
  n <- sum(rowSums(tabela))
  P <- tabela/n
  r <- rowSums(P)
  c <- colSums(P)
  d_r <- diag(r)
  d_c <- diag(c)
  R <- inv(d_r)
  C <- inv(d_c)
```

```
  A <- inv(d_r ^ (1/2)) %*% (P - r %*% t(c)) %*% inv(d_c ^ (1/2))
  total_inertia <- tr(t(A) %*% A)
  A <- svd(A)
  Gamma <- diag(A$d)
  U <- A$u
  V <- A$v
  F_ <- inv(d_r^(1/2)) %*% U %*% Gamma
  G <- inv(d_c^(1/2)) %*% V %*% Gamma
  F_ <- F_[,1:2]
  G <- G[,1:2]
  xs_row <- F_[,1] #współrzędne x dla wierszy
  ys_row <- F_[,2] #współrzędne y dla wierszy
  xs_col <- G[,1] #współrzędne x dla kolumn
  ys_col <- G[,2] #współrzędne y dla kolumn
  gam <- A$d ^ 2
  dim1 <- round(sum(gam[1])/sum(gam), 3) * 100
  dim2 <- round(sum(gam[2])/sum(gam), 3) * 100
  df_row <- data.frame('Dim.1' = xs_row, 'Dim.2' = ys_row, row.names = rownames(tabela))
  df_col <- data.frame('Dim.1' = xs_col, 'Dim.2' = ys_col, row.names = colnames(tabela))

  ggplot() + geom_point(aes(x=df_row$Dim.1, y=df_row$Dim.2), color='blue', shape = 16) +
    geom_text(aes(x=df_row$Dim.1, y=df_row$Dim.2),label=rownames(df_row),
              nudge_x = 0, nudge_y = 0.02, size=2.5, color='blue') +
    geom_point(aes(x=df_col$Dim.1, y=df_col$Dim.2), color='red', shape=17) +
    geom_text(aes(x=df_col$Dim.1, y=df_col$Dim.2),label=rownames(df_col),
              nudge_x = 0, nudge_y = -0.02, size=2.5, color='red') +
    xlab(paste('Dim 1 (', as.character(round(dim1,2)), '%)')) +
    ylab(paste('Dim 2 (', as.character(round(dim2,2)), '%)'))
}
analiza.korespondencji(tabela)
```