Rozwiązania zadań C8

Zad. 1. Z partii włókien wełny wylosowano dwie próbki i w każdej z nich zmierzono średnicę włókien różnymi metodami. W pierwszej próbce, o liczności 50, otrzymano średnią średnicę włókna: 22,9 µm; zaś w drugiej próbce, o liczności 120, otrzymano średnią średnicę włókna: 23,2 µm. Na poziomie istotności α=0,05 zweryfikować hipotezę, że w przypadku obu metod wartości oczekiwane średnicy włókna są takie same. Założyć, że wyniki pomiarów obydwiema metodami mają rozkłady normalne; pierwsza metoda charakteryzuje się znanym odchyleniem standardowym wynoszącym 4,16 µm, zaś w drugiej metodzie znane odchylenie standardowe wynosi 5,87 µm.

Rozw.

- Niech zmienna losowa X oznacza średnicę włókna zmierzoną pierwszą metodą $X \sim N(\mu_1, \sigma_1)$, przy czym $\sigma_1 = 4{,}16\ \mu\mathrm{m}$,
- Niech zmienna losowa Y oznacza średnicę włókna zmierzoną drugą metodą $Y \sim N(\mu_2, \sigma_2)$, przy czym $\sigma_2 = 5{,}87\ \mu\mathrm{m}$,
- Próbka średnic włókien otrzymanych pierwszą metodą: $n_1 = 50, \bar{x} = 22{,}9\ \mu\mathrm{m}$
- Próbka średnic włókien otrzymanych drugą metodą: $n_2 = 120, \bar{y} = 23{,}2\ \mu\mathrm{m}$
- $\alpha = 0{,}05$,
- Pytanie: zweryfikować hipotezę, że w przypadku obu metod wartości oczekiwane średnicy włókna są takie same ?

1. Hipotezy: $H_0\colon\ \mu_1 = \mu_2,\quad H_1\colon\ \mu_1 \neq \mu_2$

2. Statystyka testowa:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

   (ma rozkład standardowy normalny), jeśli hipoteza zerowa jest prawdziwa

3. Wartość statystyki testowej:

$$z = \frac{22{,}9 - 23{,}2}{\sqrt{\frac{4{,}16^2}{50} + \frac{5{,}87^2}{120}}} = -0{,}37728$$

4. Zbiór krytyczny $C = \left(-\infty, -z_{1-\frac{\alpha}{2}}\right] \cup \left[z_{1-\frac{\alpha}{2}}, \infty\right) = \left(-\infty, -z_{0{,}975}\right] \cup \left[z_{0{,}975}, \infty\right)$

   Zatem $C = (-\infty, -1{,}96] \cup [1{,}96, \infty) = \{z\colon |z| \geq 1{,}96\}$

5. Decyzja i jej uzasadnienie: $z \notin C$, więc nie ma podstaw do odrzucenia hipotezy zerowej na poziomie istotności 0,05: wyniki pomiarów nie przeczą hipotezie, że dwie metody dają równe średnie średnice włókien.

Zad. 2. Badano zmianę poziomu płac pracowników pewnego przedsiębiorstwa w latach 2001-2002. Dla 12-osobowej próby pracowników zatrudnionych w tym przedsiębiorstwie w 2001 r. otrzymano średnią płacę 1240 zł i odchylenie standardowe 110 zł, a dla 10-osobowej próby innych pracowników zatrudnionych w tym przedsiębiorstwie w 2002 r. otrzymano średnią płacę 1480 zł i odchylenie standardowe 140 zł. Zakładamy, że płace w poszczególnych latach miały rozkłady normalne o równych wariancjach. Czy na podstawie tych danych można uznać, że średnie płace w 2002 r. wzrosły w porównaniu z 2001 r.? Przyjąć poziom istotności 0,05.

Rozw.

- Niech zmienna losowa X oznacza płacę pracownika w roku 2001, $X \sim N(\mu_1, \sigma_1)$, przy czym parametry rozkładu nieznane
- Niech zmienna losowa Y oznacza płacę pracownika w roku 2002, $Y \sim N(\mu_2, \sigma_2)$, przy czym parametry rozkładu nieznane
- $\sigma_1 = \sigma_2 = \sigma$
- Próbka płac pracowników w roku 2001: liczebność $n_1 = 12$, średnia z próbki $\bar{x} = 1240$ zł., próbkowe odchylenie standardowe $s_1 = 110$ zł.
- Próbka płac pracowników w roku 2002: liczebność $n_2 = 10$, średnia z próbki $\bar{y} = 1480$ zł., próbkowe odchylenie standardowe $s_2 = 140$ zł.
- Poziom istotności: $\alpha = 0{,}05$,
- Pytanie: czy na podstawie danych zadania można uznać, że średnie pace wzrosły w roku 2002 w porównaniu z rokiem 2001?

1. Hipotezy: H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 < \mu_2$

2. Statystyka testowa:

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

(ma rozkład t-Studenta o , jeśli hipoteza zerowa jest prawdziwa, a

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

3. Wartość statystyki testowej:

$$t = \frac{1240 - 1480}{\sqrt{\frac{9 \cdot 110^2 + 11 \cdot 140^2}{20}}\sqrt{\frac{1}{12} + \frac{1}{10}}} = -4{,}50583$$

4. H_1: $\mu_1 < \mu_2 \Longrightarrow$ zbiór krytyczny $C = \left(-\infty, -t_{1-\alpha,20}\right] = \left(-\infty, -t_{0,95;20}\right]$.
Zatem $C = (-\infty, -1{,}7247] = \{t: t \leq -1{,}7247\}$

5. Decyzja i jej uzasadnienie: $t \in C$, więc odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną na poziomie istotności 0,05: wyniki pomiarów wskazują na to, że dwie metody dają różne średnie średnice włókien.

Zad.3. Badając wpływ nowego leku na poprawę stanu zdrowia chorych na cukrzycę, podano 300 losowo wybranym chorym ten nowy lek i u 240 z nich stwierdzono, po ustalonym okresie leczenia, powrót poziomu cukru w organizmie do normy. Natomiast w grupie 200 chorych leczonych lekami tradycyjnymi cukier powrócił do normy u 124 pacjentów. Na poziomie istotności 0,01 zweryfikować hipotezę, że nowy lek jest skuteczniejszy od leków tradycyjnych.

Rozw.

- Niech zmienna losowa X przyjmuje wartości 1, jeśli nowy lek poprawia stan chorego na cukrzycę, oraz 0 w przypadku przeciwnym. Niech $p_1 = P(X = 1)$, stąd

$$X \sim Bin(1, p_1),\ p_1 \in (0{,}1)$$

- Niech zmienna losowa Y przyjmuje wartości 1, jeśli tradycyjny lek poprawia stan chorego na cukrzycę, oraz 0 w przypadku przeciwnym. Niech $p_2 = P(Y = 1)$, stąd

$$Y \sim Bin(1, p_2),\ p_2 \in (0{,}1)$$

- Próbka o liczności $n_1 = 300$, $k_1 = 240$ pacjentów leczonych nowym lekiem uzyskało poprawę
- Próbka o liczebności $n_2 = 200$, $k_2 = 124$ pacjentów leczonych tradycyjnymi lekami uzyskało poprawę
- Poziom istotności: $\alpha = 0{,}01$

- **Pytanie**: zweryfikować hipotezę, że nowy lek jest skuteczniejszy od leków tradycyjnych.

Założenia zadania sugerują model 11 – test o różnicy proporcji dwóch populacji,

Jeśli spełnione są warunki:

$$n_1\hat{p}_1 = 300 \cdot \frac{240}{300} = 240 \geq 5,\ \ n_1(1-\hat{p}_1) = 300\left(1-\frac{240}{300}\right) = 60 \geq 5$$

$$n_2\hat{p}_2 = 200 \cdot \frac{124}{200} = 124 \geq 5,\ \ n_2(1-\hat{p}_2) = 200\left(1-\frac{124}{200}\right) = 76 \geq 5$$

Możemy stosować model 11.

1. Hipotezy: $H_0\colon\ p_1 = p_2,\quad H_1\colon\ \ p_1 > p_2$

2. Statystyka testowa:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{p}(1-\hat{p})}} \sim \text{ bliski } N(0{,}1),$$

ma rozkład w przybliżeniu standardowy normalny, o ile hipoteza zerowa prawdziwa oraz $n_i\hat{p}_i \geq 5,\ \ n_i(1-\hat{p}_i) \geq 5,\ \ i = 1, 2$, gdzie $\hat{p}_i = \frac{K_i}{n_i}$ oznaczają proporcje empiryczne , $\hat{p} = \frac{K_1+K_2}{n_1+n_2}$.

3. Wartość statystyki testowej:

$$z = \frac{\frac{240}{300} - \frac{124}{200}}{\sqrt{\frac{1}{300}+\frac{1}{200}} \cdot \sqrt{\frac{364}{500}\left(1-\frac{364}{500}\right)}} = \frac{0{,}8 - 0{,}62}{\sqrt{\frac{5}{600} \cdot 0{,}728}} = 4{,}431115$$

4. Zbiór krytyczny $C = [z_{1-\alpha}, \infty) = \{z\colon z \geq z_{1-\alpha}\}$, gdzie $1-\alpha = 0{,}99$, $z_{0,99} = 2{,}326348$, zatem

$$C = [2{,}326348, \infty) = \{z\colon z \geq 2{,}326348\}$$

5. Decyzja i jej uzasadnienie: $z \in C$, więc odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną na poziomie istotności 0,01: efekty leczenia wskazują na to, że nowy lek jest skuteczniejszy od tradycyjnych leków.

<u>Zad.4</u>. Dwie formacje geologiczne porównano pod względem zawartości pewnego minerału. Uzyskano następujące dane:

Formacja I	7,6	11,1	6,8	9,8	4,9	6,1	15,1
Formacja II	4,7	6,4	4,1	3,7	3,9		

Zakładamy, że rozkłady zawartości tego minerału w obu formacjach są normalne z odchyleniami standardowymi równymi odpowiednio: 2 i 1. Czy można stwierdzić, że średnia zawartość tego minerału w pierwszej formacji jest istotnie większa od zawartości w drugiej formacji? Przyjąć poziom istotności 0,05.

Rozw.

- Niech zmienna losowa X oznacza zawartość minerału w formacji I, $X \sim N(\mu_1, \sigma_1)$, przy czym $\sigma_1 = 2$, średnia nieznana
- Niech zmienna losowa Y oznacza zawartość minerału w formacji II , $Y \sim N(\mu_2, \sigma_2)$, przy czym $\sigma_2 = 1$, średnia nieznana

- Próbka dla formacji I: 7,6 11,1 6,8 9,8 4,9 6,1 15,1

  Stąd liczebność $n_1 = 7$, średnia z próbki $\bar{x} = \frac{61,4}{7} \cong 8,77143$

- Próbka dla formacji II: 4,7 6,4 4,1 3,7 3,9

  Stąd liczebność $n_2 = 5$, średnia z próbki $\bar{y} = \frac{22,8}{5} = 4,56$

- Poziom istotności: $\alpha = 0,05$
- Pytanie: czy można twierdzić, ze średnia zawartość minerału w formacji I jest większa niż w formacji II?

1. Hipotezy: H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 > \mu_2$

2. Statystyka testowa:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1),$$

jeśli hipoteza zerowa jest prawdziwa

3. Wartość statystyki testowej:

$$z = \frac{8,77143 - 4,56}{\sqrt{\frac{4}{7} + \frac{1}{5}}} = 4,79492$$

4. Zbiór krytyczny $C = [z_{1-\alpha}, \infty) = \{z : z \geq z_{1-\alpha}\}$, gdzie $1 - \alpha = 0,95$,

   $z_{0,95} = 1,644854$, zatem

$$C = [1{,}644854; \infty) = \{z: z \geq 1{,}644854\}$$

5. Decyzja i jej uzasadnienie: $z \in C$, więc odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną na poziomie istotności 0,05: wyniki pomiarów wskazują na to, średnia zawartość tego minerału w pierwszej formacji jest większa od zawartości w drugiej na poziomie istotności 0,05.

Uwaga: Możemy jeszcze znaleźć p-wartość w celu określenia jak „silne" jest nasze przekonanie, ze hipoteza alternatywna jest prawdziwa:

p-wartość = $P(Z \geq 4{,}79492) = 1 - 0{,}9999991 \ldots \cong 0$ = najmniejszy poziom istotności prowadzący do odrzucenia hipotezy zerowej, czyli przyjmując poziom istotności prawie 0 prawdop. błędu I rodzaju (przyjęcia hipotezy alternatywnej fałszywej wynosi 0)

<u>Zad.5</u>. Wykonano 100 rzutów sześciościenną kostką do gry i otrzymano następujące liczebności wyników:

Liczba oczek	1	2	3	4	5	6
Liczba wystąpień	16	19	9	17	25	14

Czy istnieją podstawy do odrzucenia hipotezy, że rzuty były wykonywane uczciwą kostką, czyli do odrzucenia hipotezy o jednostajności rozkładu liczby wyrzuconych oczek? Przyjąć poziom istotności 0,05.

<u>Rozw.</u>

- X – liczba wyrzuconych oczek w rzucie kostką sześcienną
- $P(X = j) = p_j,\ j = 1,2,3,4,5,6$, gdzie $p_j \in (0,1)$ oraz

$$\sum_{j=1}^{6} p_j = 1.$$

- Poziom istotności: $\alpha = 0{,}05$
- Pytanie: czy można twierdzić, że kostka jest uczciwa, czyli rozkład liczby wyrzuconych oczek jest jednostajny na zbiorze {1,2,3,4,5,6}?

1. Hipotezy: H_0: $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$, H_1: istnieje j_0: $p_{j_0} \neq \frac{1}{6}$

2. Statystyka testowa:

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2_{k-1},$$

gdzie $k = 6$ jest liczbą wartości zmiennej losowej X, a N_i jest liczbą rzutów (elementów próby losowej o wartości i), w których wystąpiła liczba oczek, $i = 1.2,3,4,5,6$.

$n = N_1 + N_2 + N_3 + N_4 + N_5 + N_6 = 100$ jest liczbą rzutów (liczebnością próby losowej, której

3. Wartość statystyki testowej

$$\chi^2_{obs} = 8{,}4800$$

Nr i	Klasa c_i	n_i	p_i	$np_i = \frac{100}{6} = \frac{50}{3}$	$\frac{\left(n_i - \frac{50}{3}\right)^2}{\frac{50}{3}}$
1	1	16	1/6	50/3	0,02677
2	2	19	1/6	50/3	0,32667
3	3	9	1/6	50/3	3,52667
4	4	17	1/6	50/3	0,00667
5	5	25	1/6	50/3	4,16667
6	6	14	1/6	50/3	0,42667

4. Zbiór krytyczny $C = \left[\chi^2_{0,95;5}; \infty\right) = \left\{\chi^2 : \chi^2 \geq \chi^2_{095;5}\right\}$, gdzie $1 - \alpha = 0{,}95$,

$\chi^2_{095;5} = 11{,}07$, zatem

$$C = [11{,}07; \infty)$$

5. Decyzja i jej uzasadnienie:

$$\chi^2_{obs} = 8{,}4800 \notin C,$$

więc nie ma podstaw do odrzucenia hipotezy zerowej, ze kostka jest uczciwa na poziomie istotności 0,05.

<u>Zad.6</u>. W klasycznych doświadczeniach dotyczących selekcji grochu Mendel obserwował liczności występowania różnych rodzajów nasion otrzymanych przy krzyżowaniu roślin z okrągłymi i żółtymi nasionami oraz roślin z pomarszczonymi i zielonymi nasionami. Otrzymał następujące wyniki:

pomarszczone i zielone	32
okrągłe i zielone	108
pomarszczone i żółte	101
okrągłe i żółte	315

Według teoretycznych rozważań prawdopodobieństwa występowania wymienionych rodzajów nasion winny być w stosunku 1 : 3 : 3 : 9. Na poziomie istotności 0,05 zweryfikować hipotezę, że rozkład prawdopodobieństwa liczby wymienionych czterech rodzajów nasion jest zgodny z teorią.

Rozw.

- X – rodzaj otrzymanego nasiona w wyniku krzyżowania
- Wyniki eksperymentu:

Nr klasy i	Klasa (rodzaj) c_i	Liczność n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
1	pomarszczone i zielone	32			
2	okrągłe i zielone	108			
3	pomarszczone i żółte	101			
4	okrągłe i żółte	315			

- Pytanie: czy można twierdzić, że rozkład rodzajów nasion X występuje w stosunku: 1;3;3:9?
- Poziom istotności: $\alpha = 0{,}05$

1. Hipotezy: H_0: $p_1 = \frac{1}{16}$, $p_2 = \frac{3}{16}$, $p_3 = \frac{3}{16}$, $p_4 = \frac{9}{16}$, H_1: zaprzeczenie H_0

2. Statystyka testowa:

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2_{k-1},$$

gdzie $k = 4$ jest liczbą wartości zmiennej losowej X, a N_i jest liczbą ziaren rodzaju $i -$ go (elementów próby losowej o wartości i), w próbce o liczności $n = 556$.

$n = N_1 + N_2 + N_3 + N_4 = 556$

3. Wartość statystyki testowej

$$\chi^2_{obs} = 0,4700$$

Nr klasy i	Klasa (rodzaj) c_i	Liczność n_i	H_0 p_i	np_i $n = 556$	$\frac{(n_i - np_i)^2}{np_i}$
1	pomarszczone i zielone	32	1/16	34,75	0,21763
2	okrągłe i zielone	108	3/16	104,25	0,13489
3	pomarszczone i żółte	101	3/16	104,25	0,10132
4	okrągłe i żółte	315	9/16	312,75	0,01619
			Suma		0,4700

4. Zbiór krytyczny $C = \left[\chi^2_{095;3}, \infty\right) = \left\{\chi^2 : \chi^2 \geq \chi^2_{095;3}\right\}$, gdzie $1 - \alpha = 0,95$,

$\chi^2_{095;3} = 7,814728$, zatem

$$C = [7,814728; \infty)$$

5. Decyzja i jej uzasadnienie:

$$\chi^2_{obs} = 0,4700 \notin C,$$

więc nie ma podstaw do odrzucenia hipotezy zerowej, że rozkład prawdopodobieństwa rodzaju ziarna jest taki jak podaje H_0 na poziomie istotności 0,05.

Zad.7. Badano zależność między liczbą wypalanych papierosów a wystąpieniem pewnych niekorzystnych zmian w płucach w grupie 1500 osób. Zebrane dane przedstawiono w tabeli kontyngencyjnej:

	niepalący	palący mało	palący dużo
zmiany występują	51	250	560
zmian nie ma	370	210	59

Zweryfikować zależność między cechami, przyjmując poziom istotności α=0,01.

Rozw.

- (X,Y)
- Pytanie: X, Y są niezależne?
- Poziom istotności: $\alpha = 0,01$
-

- Wyniki eksperymentu:

Y X	Niepalący 1	palący mało 2	palący dużo 3
zmiany występują = 1	51	250	560
zmian nie ma = 2	370	210	59

$n = \sum_{i=1}^{2} \sum_{j=1}^{3} n_{ij} = 1500$ – liczebność próbki badanej cechy dwuwymiarowej

1. Hipotezy: H_0: X, Y są niezależnymi zmiennymi losowymi,

   H_1: zaprzeczenie H_0, tzn. zmienne losowe X, Y są zależne

2. Statystyka testowa:

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{\left(N_{ij} - \widehat{N_{ij}}\right)^2}{\widehat{N_{ij}}} \sim \chi_2^2, \qquad \text{jeśli hipoteza zerowa prawdziwa}$$

Liczba stopni swobody = (2 – 1)(3-1) = 2, $\widehat{N_{ij}} = \frac{N_{i\cdot} \cdot N_{\cdot j}}{n}$ = estymator $E(N_{ij})$

$N_{i\cdot} = \sum_{j=1}^{3} N_{ij}$ = liczba elementów próby losowej, dla których cecha X ma wartość i-tą

$N_{\cdot j} = \sum_{i=1}^{2} N_{ij}$ = liczba elementów próby losowej, dla których cecha Y ma wartość j-tą

3. Wartość statystyki testowej

Y X	Niepalący 1	palący mało 2	palący dużo 3	$n_{i\cdot}$
zmiany występują = 1	51 241,65	250 264,04	560 355,31	861
zmian nie ma = 2	370 179,35	210 195,96	59 263,69	639
$n_{\cdot j}$	421	460	619	1500

gdzie w tabeli policzono: $\widehat{n_{ij}} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$, $np \quad \widehat{n_{11}} = \frac{n_{1\cdot} \cdot n_{\cdot 1}}{n} = \frac{861 \cdot 421}{1500} = 241,65$

$$\chi^2_{obs.} = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{\left(n_{ij}-\widehat{n_{\iota J}}\right)^2}{\widehat{n_{\iota J}}} =$$

$$\frac{(51-861\cdot 421/1500)^2}{861\cdot 421/1500}+\frac{(250-861\cdot 460/1500)^2}{861\cdot 460/1500}+\frac{(560-861\cdot 619/1500)^2}{861\cdot 619/1500}$$

$$\frac{(370-639\cdot 421/1500)^2}{639\cdot 421/1500}+\frac{(210-639\cdot 460/1500)^2}{639\cdot 460/1500}+\frac{(59-639\cdot 619/1500)^2}{639\cdot 619/1500}$$

$$\chi^2_{obs.} = 631{,}665$$

4. Zbiór krytyczny $C = \left[\chi^2_{0,99;2}, \infty\right) = \left\{\chi^2 : \chi^2 \geq \chi^2_{0,99;2}\right\}$, gdzie $1-\alpha = 0{,}99$,

   $\chi^2_{0,99;2} = 9{,}21034$, zatem
$$C = [9{,}21034; \infty)$$

5. Decyzja i jej uzasadnienie:
$$\chi^2_{obs} = 631{,}665 \in C,$$

   więc należy odrzucić hipotezę zerową, że cechy są niezależne. Można twierdzić, że X, Y są zależnymi zmiennymi losowymi (palenie papierosów i zmiany w płucach są zależne) na poziomie istotności 0,05.

Zad.8. W próbie liczącej 100 mężczyzn w wieku 50-60 lat zbadano częstość występowania choroby wieńcowej i podwyższonego ciśnienia tętniczego. Zebrane dane przedstawiono w tabeli kontyngencyjnej:

	ciśnienie niepodwyższone	ciśnienie podwyższone
choroba wieńcowa nie występuje	37	17
choroba wieńcowa występuje	8	38

Na poziomie istotności 0,01 ocenić, czy choroba wieńcowa współistnieje z podwyższonymi wartościami ciśnienia tętniczego.

Rozw.

- (X,Y)
- Pytanie: X, Y są niezależne?
- Poziom istotności: $\alpha = 0{,}01$

- Wyniki eksperymentu:

Y X	Ciśnienie niepodwyższone = 1	ciśnienie podwyższone = 0	$n_{i\cdot}$
choroba wieńcowa nie występuje = 1	37	17	54
choroba wieńcowa występuje = 0	8	38	46
$n_{\cdot j}$	45	55	100

1. Hipotezy:

- H_0: X, Y są niezależnymi zmiennymi losowymi, występowanie bądź nie choroby wieńcowej oraz wysokość ciśnienia są cechami niezależnymi
- H_1: zaprzeczenie H_0, tzn. zmienne losowe X, Y są zależne

2. Statystyka testowa:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(N_{ij} - \widehat{N_{ij}}\right)^2}{\widehat{N_{ij}}} \sim \chi_1^2, \quad \text{jeśli } X, Y \text{ są niezależne}$$

Liczba stopni swobody = (2 – 1)(2 - 1) = 1, $\widehat{N_{ij}} = \frac{N_{i\cdot} \cdot N_{\cdot j}}{n}$ = estymator $E\left(N_{ij}\right)$

3. Wartość statystyki testowej

Y / X	Ciśnienie niepodwyższone = 1	ciśnienie podwyższone = 0	$n_{i\cdot}$
choroba wieńcowa nie występuje = 1	37 24,3	17 29,7	54
choroba wieńcowa występuje = 0	8 20,7	38 25,3	46
$n_{\cdot j}$	45	55	100

gdzie w tabeli policzono: $\widehat{n_{ij}} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$, np $\widehat{n_{11}} = \frac{n_{1\cdot} \cdot n_{\cdot 1}}{n} = \frac{54 \cdot 45}{100} = 24{,}3$

$$\chi^2_{obs.} = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(n_{ij} - \widehat{n_{ij}}\right)^2}{\widehat{n_{ij}}} = \frac{(37-24{,}3)^2}{24{,}3} + \frac{(17-29{,}7)^2}{29{,}7} + \frac{(8-20{,}7)^3}{20{,}7} + \frac{(38-25{,}3)^2}{25{,}3}$$

$$\chi^2_{obs.} = 26{,}2350$$

4. Zbiór krytyczny $C = \left[\chi^2_{0,99;1}, \infty\right) = \left\{\chi^2 : \chi^2 \geq \chi^2_{0,99;1}\right\}$, gdzie $1 - \alpha = 0{,}99$,

$\chi^2_{0,99;1} = 6{,}6349$, zatem

$$C = [6{,}6349; \infty)$$

5. Decyzja i jej uzasadnienie:

$$\chi^2_{obs} = 26{,}2350 \in C$$

Można twierdzić, że X, Y są zależnymi zmiennymi losowymi, przy poziomie istotności 0,01.