

3

Lista 3

Część I i II

Zadanie 1

Napisz funkcję, która zwraca p-wartość w omówionym na wykładzie warunkowym teście symetrii w przypadku tabeli 2×2 .

```
p <- function(n12, n21){  
  part <- 0  
  if(n12 < (n12+n21)/2){  
    for(i in 0:n12){  
      part <- part + choose(n12+n21,i)*(1/2)^i*(1/2)^(n12+n21-i)  
    }  
    part <- 2*part  
  }  
  if(n12 > (n12+n21)/2){  
    for(i in 0:n12+n21){  
      part <- part + choose(n12+n21,i)*(1/2)^i*(1/2)^(n12+n21-i)  
    }  
    part <- 2*part  
  }  
  if(n12 == (n12+n21)/2){  
    part <- 1  
  }  
  return(part)  
}
```

Zadanie 2

Zadanie 2.1

```
tabela <- matrix(c(1, 2, 5, 4), nrow = 2,  
dimnames = list("Lek A" = c("Negatywna", "Pozytywna"),  
"Lek B" = c("Negatywna", "Pozytywna")))  
print(tabela)
```

	Lek B	
Lek A	Negatywna	Pozytywna
Negatywna	1	5
Pozytywna	2	4

```
#test McNemara  
mcnemar.test(tabela, correct = TRUE)
```

McNemar's Chi-squared test with continuity correction

data: tabela

McNemar's chi-squared = 0.57143, df = 1, p-value = 0.4497

Zadanie 2.2

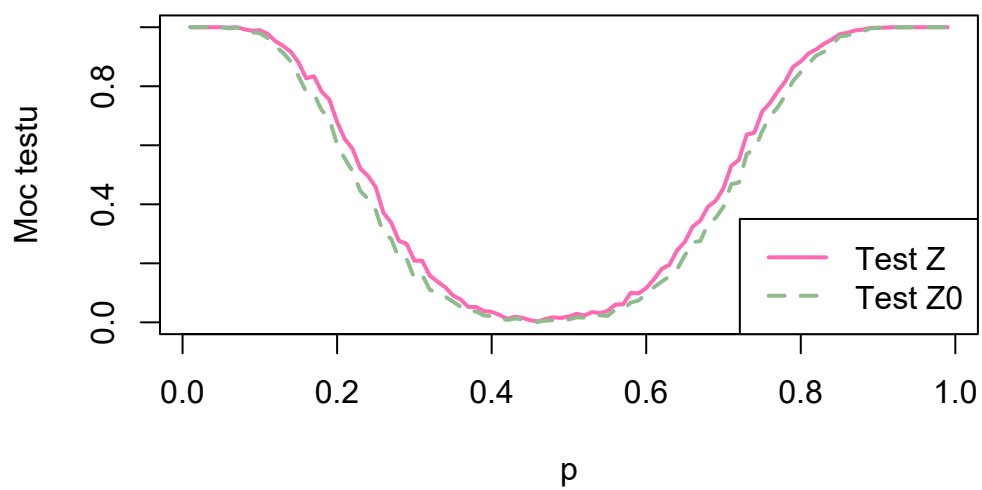
[1] 0.453125

[1] 0.453125

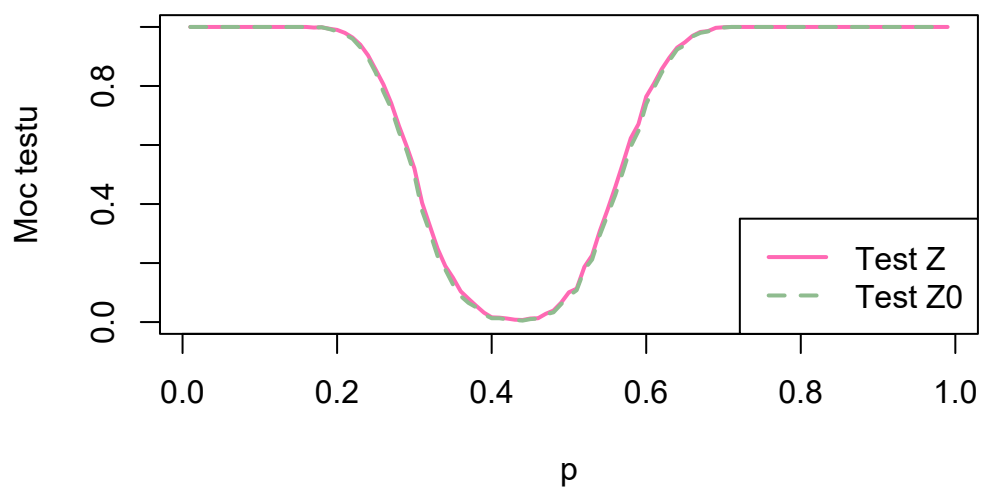
Zadanie 3

Przeprowadź symulacje w celu porównania mocy testu Z i testu Z_0 przedstawionych na wykładzie. Rozważ różne długości prób.

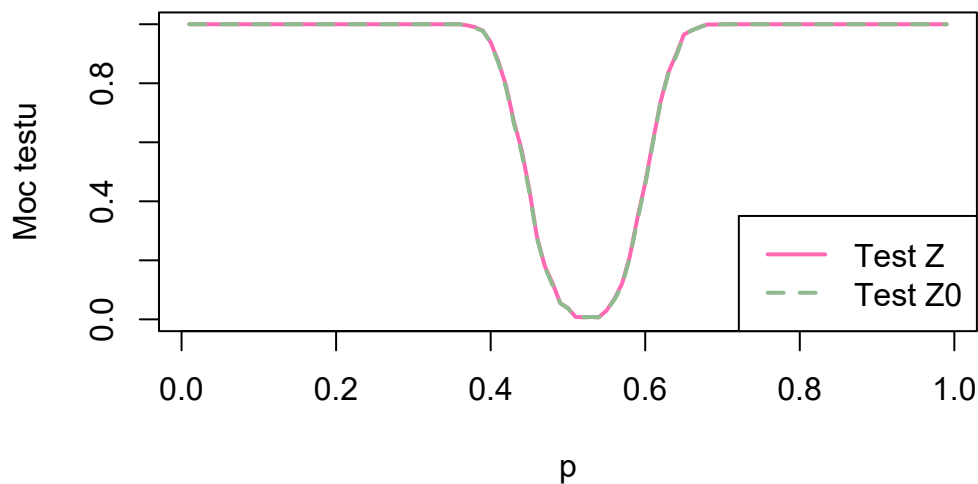
Wykres mocy testów Z i Z0 dla n=30



Wykres mocy testów Z i Z0 dla n=100



Wykres mocy testów Z i Z0 dla n=300



Widzimy, że dla mniejszych n test Z ma większą moc od testu Z_0 . Dla większych n moce testów zbliżają się do siebie oraz rosną, szczególnie wokół $p = 0.5$.

Zadanie 4

McNemar's Chi-squared test with continuity correction

data: tabela

McNemar's chi-squared = 4.3214, df = 1, p-value = 0.03764

Część III

Zadanie 6

W pewnym badaniu porównywano skuteczność dwóch metod leczenia: Leczenie A to nowa procedura, a Leczenie B to stara procedura. Przeanalizuj dane przedstawione w Tabeli 3 (wyniki dla całej grupy pacjentów) oraz w Tabelach 4 i 5 (wyniki w podgrupach ze względu na dodatkową zmienną) i odpowiedz na pytanie, czy dla danych występuje paradoks Simpsona.

Table 1: Tabela 3: Dane dla całej grupy

Metoda	Poprawa	Brak poprawy
Leczenie A	117	104
Leczenie B	177	44

Table 2: Tabela 4: Dane dla pacjentów z chorobami współistniejącymi.

Metoda	Poprawa	Brak poprawy
Leczenie A	17	101
Leczenie B	2	36

Table 3: Tabela 5: Dane dla pacjentów bez chorób współistniejących.

Metoda	Poprawa	Brak poprawy
Leczenie A	100	3
Leczenie B	175	8

```

wszyscy z chorobami bez chorób
A 0.5294118 0.14406780 0.9708738
B 0.8009050 0.05263158 0.9562842

```

Chociaż leczenie B “wygrywa” patrząc na całą grupę badanych, po podziale na grupy ze względu na obecność chorób współistniejących możemy zauważyć, że to leczenie A ma większy odsetek wyzdrowień.

```

wszyscy z chorobami bez chorób
2.740007e-09 2.248419e-01 7.675118e-01

```

W przeprowadzonym teście niezależności χ^2 dla całej grupy p-value jest bardzo małe, więc odrzucamy hipotezę H_0 o niezależności. Jednak ten sam test wykonany osobno dla badanych grup - z chorobami współistniejącymi oraz bez chorób - w obu przypadkach daje p-value większą od poziomu istotności, a więc nie mamy podstaw do odrzucania hipotezy zerowej o niezależności zmiennych, to znaczy wyniku leczenia (poprawy) od przyjętego leczenia. To znaczy, że pozorny związek dla całej badanej grupy nie przekłada się na zależność w podgrupach - a więc jest to klasyczny przypadek paradoksu Simpsona.

Zadanie 7

Dla danych z listy 1, przyjmując za zmienną 1 zmienną CZY_KIER, za zmienną 2 – zmienną PYT_2 i za zmienną 3 – zmienną STAŻ, podaj interpretacje następujących modeli log-liniowych: [1 3], [13], [1 2 3], [12 3], [12 13] oraz [1 23].

[1 3] - zmienne CZY_KIER oraz STAŻ są niezależne,

[13] - zmienne CZY_KIER oraz STAŻ nie są niezależne,

[1 2 3] - zmienne CZY_KIER, PYT_2 oraz STAŻ są niezależne,

[12 3] - zmienne CZY_KIER i PYT_2 nie są niezależne, a zmienna STAŻ jest niezależna od nich obu,

[12 13] - zmienne CZY_KIER i PYT_2 nie są niezależne, CZY_KIER i STAŻ nie są niezależne, a PYT_2 i STAŻ są warunkowo niezależne,

[1 23] - zmienna CZY_KIER jest niezależna od pozostałych dwóch, PYT_2 i STAŻ, które nie są od siebie niezależne.

Część IV i V

Zadanie 8

Przyjmując model log-liniowy [123] dla zmiennych opisanych w zadaniu 7 oszacuj prawdopodobieństwa:

- że osoba pracująca na stanowisku kierowniczym jest zdecydowanie zadowolona ze szkoleń;
- że osoba o stażu pracy krótszym niż rok pracuje na stanowisku kierowniczym;
- że osoba o stażu pracy powyżej trzech lat nie pracuje na stanowisku kierowniczym.

Jakie byłyby oszacowania powyższych prawdopodobieństw przy założeniu modelu [12 23]?

```
# A tibble: 4 x 5
  PYT_2 freq_sum fitted_sum p_dane p_model
  <fct>   <int>      <dbl> <dbl>   <dbl>
1 -2         10      10.0  0.370  0.370
2 -1          2       2.00  0.0741 0.0741
3 1           2       2.00  0.0741 0.0741
4 2          13      13.0  0.481  0.481
```

```
# A tibble: 1 x 5
  STAŽ freq_sum fitted_sum p_dane p_model
<fct>   <int>      <dbl> <dbl>   <dbl>
1 1         41      41.0     1       1
```

```
# A tibble: 1 x 5
  STAŽ freq_sum fitted_sum p_dane p_model
<fct>   <int>      <dbl> <dbl>   <dbl>
1 3         19      19.0     1       1
```