Zadanie 5

Porównywano skuteczność (...).

Populacja: Pacjenci
Cecha zależna: Stan zdrowia po terapii (Choroba ustąpiła/Choroba nie ustąpiła)
Cecha niezależna: Rodzaj terapii (Terapia 1/Terapia 2)

1. Ustalone są wartości w wierszu *Ogółem*, ponieważ liczba pacjentów przydzielonych do eksperymentu jest ustalona dla każdej terapii. Pozostałe wartości są realizacjami zmiennych losowych.
2. W eksperymencie ustalono z góry, że 50% osób zostanie poddanych terapii pierwszej i 50% drugiej. W całej populacji pacjentów tak być nie musi. Te ustalone 50% nie jest żadnym oszacowaniem. Jest to tylko i wyłącznie wartość ustalona z góry.
3. Tak. Jeżeli przyjmiemy, że X_1 jest zmienną losową oznaczającą liczbą pacjentów, która może zostać wyleczona za pomocą pierwszej terapii, to możemy przyjąć, że $X_1 \sim Bin(36, \theta_1)$ (powiemy: X_1 ma rozkład dwumianowy o parametrach 36 oraz θ_1. Parametr θ_1 może być interpretowany, jako skuteczność terapii pierwszej. Możemy też powiedzieć, że $\theta_1 \cdot 100\%$ jest odsetkiem pacjentów leczonych za pomocą terapii pierwszej, którzy przychodzą do zdrowia.)
4. Odsetek pacjentów leczonych za pomocą terapii pierwszej, którzy przychodzą do zdrowia, można oszacować za pomocą przedziału ufności dla parametru θ_1.
5. Możemy zdefiniować X_2 jako zmienną losową oznaczającą liczbą pacjentów, która może zostać wyleczona za pomocą drugiej terapii i przyjąć, że $X_2 \sim Bin(36, \theta_2)$. Zatem, jeżeli oszacujemy iloraz θ_1/θ_2 za pomocą przedziału ufności, to tym samym będziemy mogli określić, ile razy skuteczniejsza jest terapia pierwsza od drugiej.
6. Stosunek liczby pacjentów leczonych skutecznie do liczby pacjentów leczonych nieskutecznie możemy dla terapi pierwszej wyrazić przez iloraz $\theta_1/(1-\theta_1)$, a dla drugiej terapii przez iloraz $\theta_2/(1-\theta_2)$. Tego typu ilorazy można interpretować jako szansę na wyleczenie odpowiednio przy terapii pierwszej i drugiej. Dlatego iloraz
$$\frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)}$$
nazywamy ilorazem szans. W tym punkcie chodzi właśnie o oszacowanie ilorazu szans:
$$OR(\theta_1|\theta_2) = \frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)}.$$
Skrót OR pochodzi od terminu **Odds Ratio.** Podany iloraz szans $OR(\theta_1|\theta_2)$ zapiszę symbolicznie słowami w następujący sposób:
$$\text{ustąpiła:nie ustąpiła/Tearapia 1:Terapia 2.}$$
lub równoważnie
$$\text{Tearapia 1:Terapia 2/ustąpiła:nie ustąpiła.}$$

Przedział ufności dla θ_1 – obliczenia w R. Do wyznaczenia przedziału ufności dla θ_1 możemy posłużyć się komendą:

```
prop.test(18,36)
#lub
binom.test(18,36)
```

Iloraz szans – Obliczenia w R. Do wyznaczenia ilorazu szans w pakiecie **R** można posłużyć się biblioteką **vcd** i użyć funkcji `oddsratio`. W tym celu należy wczytać tę bibliotekę oraz przygotować dane:

```
library(vcd)
dane=c(18,25,18,11)
dane=matrix(dane,ncol=2,byrow = T)
colnames(dane)=c("Tearapia 1","Terapia 2")
rownames(dane)=c("nie ustąpiła","ustąpiła")
names(dimnames(dane))=c("Choroba","Terapia")
```

W zależności od podejścia do problemu iloraz szans można definiować inaczej, niż to zostało przedstawione w punkcie 6. Inne warianty, to

1. $\frac{\theta_2/(1-\theta_2)}{\theta_1/(1-\theta_2)}$ – ustąpiła:nie ustąpiła/Terapia 2:Tearapia 1

   `exp(confint(oddsratio(dane[2:1,2:1])))`

2. $\frac{(1-\theta_1)/\theta_1}{(1-\theta_2)/\theta_2}$ – nie ustąpiła:ustąpiła/Tearapia 1:Terapia 2

   `exp(confint(oddsratio(dane)))`

3. $\frac{(1-\theta_2)/\theta_2}{(1-\theta_1)/\theta_1}$ – nie ustąpiła:ustąpiła/Terapia 2:Tearapia 1

   `exp(confint(oddsratio(dane[,2:1])))`

Przy obliczaniu ilorazu szans należy zwrócić uwagę na to, który z ilorazów chcemy wyznaczyć. Jeżeli chcę oszacować

$$OR(\theta_1|\theta_2) = \frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)},$$

to ze względu na sposób przygotowania danych w tym zadaniu, powinno się użyć komendy

`exp(confint(oddsratio(dane[2:1,])))`,

która daje wydruk:

```
                                                 2.5 %   97.5 %
ustąpiła:nie ustąpiła/Tearapia 1:Terapia 2 0.8665509 5.960746
```

Stąd przedział ufności, na poziomie ufności 0.95, dla $OR(\theta_1|\theta_2)$ jest postaci

$$(0.8665509, 5.960746).$$

Do obliczenia ilorazu szans można posłużyć się również biblioteką **epitools** oraz funkcją `odsratio`. Ze względu na tę samą nazwę funkcji w dwóch różnych bibliotekach, wywołanie funkcji `odsratio` powinno się odbywać następująco: `epitools::odsratio` lub `vcd::odsratio` w przypadku, gdy obie biblioteki są aktywne (można też jedną z nich dezaktywować, np `detach("package:vcd", unload=TRUE)`). Użycie funkcji:

```
oddsratio(t(dane),rev="r")
```

daje nam wydruk

```
$`data`
Choroba
Terapia      nie ustąpiła ustąpiła Total
Terapia 2            25       11    36
Tearapia 1           18       18    36
Total                43       29    72

$measure
odds ratio with 95% C.I.
Terapia      estimate     lower    upper
Terapia 2  1.000000        NA       NA
Tearapia 1 2.236739 0.8553744 6.070718

$p.value
two-sided
Terapia      midp.exact fisher.exact chi.square
Terapia 2          NA           NA         NA
Tearapia 1   0.101275    0.1488388 0.09256503

$correction
[1] FALSE

attr(,"method")
[1] "median-unbiased estimate & mid-p exact CI"
Możemy wydruk ten ograniczyć poprzez \texttt{oddsratio(t(dane),rev="r")\$measure}
```

Iloraz prawdopodopodobieństw – Obliczenia w R. Do wyznaczenia przedziału ufności możemy zastosować funkcję z biblioteki **epitools**. Iloraz θ_1/θ_2 możemy oszacować następująco:

```
riskratio(t(dane),rev="r")$measure
```

Jak zorganizować dane (pod pakiet epitools!), aby bez dodatkowych zabiegów zamiany kolumn lub wierszy wyznaczyć od razu to, co chcemy? W tym celu rozważmy następującą organizację danych:

$Y-$ cecha objaśniana (zależna)
$X-$ cecha objaśniająca (niezależna)

	Y	
X	2	1
2	$1-\theta_2$	$1-\theta_1$
1	θ_2	θ_1

W takim przypadku bezpośrednie użycie procedury `odsratio` daje nam oszacowanie ilorazu szans postaci $OR(\theta_1|\theta_2) = \frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)}$, a użycie `riskratio` oszacowanie ilorazu $RR(\theta_1|\theta_2) = \frac{\theta_1}{\theta_2}$.

Dlatego odpowiednio organizując dane możemy oszacować interesujące nas wielkości bez dodatkowych zabiegów:

```
> dane=matrix(c(25,18,11,18),ncol=2)
> colnames(dane)=c("nie ustąpiła","ustąpiła")
> rownames(dane)=c("Terapia 2","Terapia 1")
> names(dimnames(dane))=c("Terapia","Choroba")
> dane
            Choroba
Terapia      nie ustąpiła ustąpiła
Terapia 2            25       11
Terapia 1            18       18

>epitools::oddsratio(dane)$measure
odds ratio with 95% C.I.
Terapia      estimate     lower    upper
Terapia 2  1.000000        NA       NA
Terapia 1  2.236739 0.8553744 6.070718

> riskratio(dane)$measure
risk ratio with 95% C.I.
Terapia      estimate     lower    upper
Terapia 2  1.000000        NA       NA
Terapia 1  1.636364 0.9062184 2.954791
>
```

Wnioski:

1. $\widehat{OR}(\theta_1|\theta_2) = 2.236739$ oraz $OR(\theta_1|\theta_2) \in (0.8553744, 6.070718)$
2. $\widehat{RR}(\theta_1|\theta_2) = 1.636364$ oraz $RR(\theta_1|\theta_2) \in (0.9062184, 2.954791)$

Zobaczmy teraz, co zostanie oszacowane przez instrukcję `oddsratio` z biblioteki **vcd** przy tak zorganizowanych danych:

```
> exp(confint(vcd::oddsratio(dane)))
                                                  2.5 %   97.5 %
Tearapia 2:Terapia 1/nie ustąpiła:ustąpiła 0.8665509 5.960746
```

Wydruk sugeruje, że został oszacowany iloraz szans:

$$OR(1-\theta_2|1-\theta_1) = \frac{(1-\theta_2)/\theta_2}{(1-\theta_1)/\theta_1}.$$

Zauważmy jednak, że

$$OR(1-\theta_2|1-\theta_1) = \frac{(1-\theta_2)/\theta_2}{(1-\theta_1)/\theta_1} = \frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)} = OR(\theta_1|\theta_2),$$

co wyjaśnia, dlaczego oszacowanie jest identyczne jak w przypadku:

```
> exp(confint(vcd::oddsratio(dane[2:1,2:1])))
2.5 %   97.5 %
Terapia 1:Tearapia 2/ustąpiła:nie ustąpiła 0.8665509 5.960746
```

dla danych w postaci

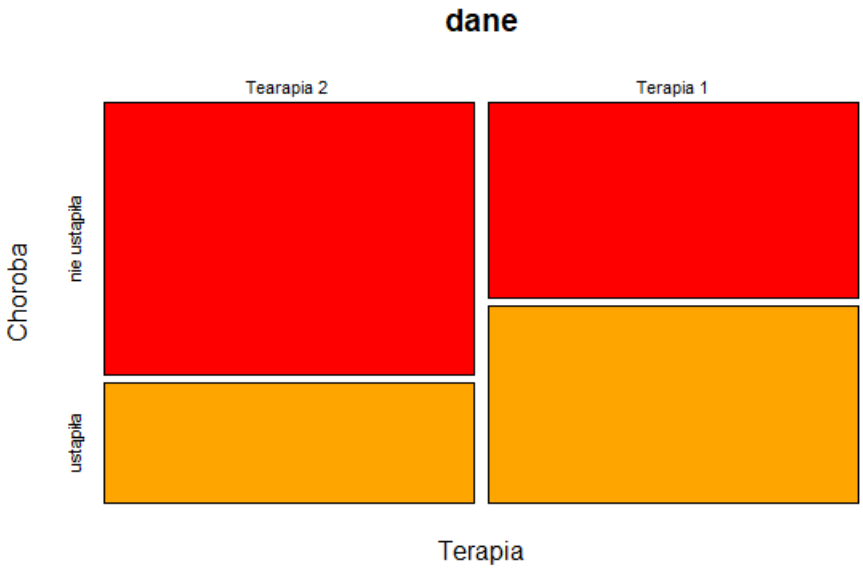
```
> dane[2:1,2:1]
           Choroba
Terapia      ustąpiła nie ustąpiła
Terapia 1        18           18
Tearapia 2       11           25
```

Uwaga: Oszacowanie ilorazu szans za pomocą procedury `oddsratio` zależy od biblioteki: (0.85537446.070718) w bibliotece **epitools** oraz $(0.8665509, 5.960746)$ w **vcd**. Różnica jest niewielka. W obu przypadkach przedział ufności jest przybliżony.

Wnioski: Oszacowanie punktowe ilorazu szans $OR(\theta_1|\theta_2)$ równe 2.236739 sugeruje, że szansa na wyleczenie przy terapii pierwszej jest ok. 2.3 razy większa niż przy terapii drugiej. Nie jest to jednak wynik statystycznie istotny. Przedział ufności jest postaci (0.85537446.070718), co oznacza, że $OR(\theta_1|\theta_2)$ może być równie dobrze mniejszy od 1, jak i większy. Zatem musimy uznać, że z danych nie wynika, aby któraś z terapii była lepsza.

Ilustracja graficzna

```
> mosaicplot(dane,col=c("red","orange"))
```


dane
Tearapia 2
Terapia 1
Choroba
nie ustąpiła
ustąpiła
Terapia