Raport 1

Zofia Górnikowska, Michał Wiktorowski

Marzec 2023

1 Wstęp

W sprawozdaniu poddamy analizie dane nierzeczywiste, wygenerowane na potrzeby zajęć laboratoryjnych. Zakładamy, że zostały uzyskane w wyniku badań ankietowych losowo wybranych dwustu pracowników pewnej wielkiej korporacji. Pytania ankietowe były następujące (w nawiasach podane są przyjęte nazwy zmiennych i symbole kodowania odpowiedzi w zbiorze danych).

1. (D) Pracuję
   (a) w dziale zaopatrzenia (Z),
   (b) w dziale produkcyjnym (P),
   (c) w dziale sprzedaży (w tym marketingu) (S),
   (d) w dziale obsługi kadrowo-płacowej (O).
2. (S) Pracuję na stanowisku kierowniczym (tzn. kieruję pracą więcej niż pięciu osób)
   (a) tak (1),
   (b) nie (0).
3. (A1 i A2) Atmosfera w miejscu pracy jest bardzo dobra
   (a) zdecydowanie się nie zgadzam (-2),
   (b) nie zgadzam się (-1),
   (c) trudno powiedzieć (0),
   (d) zgadzam się (1),
   (e) zdecydowanie się zgadzam (2).
4. (W1 i W2) Jestem zadowolona/y ze swojego wynagrodzenia
   (a) zdecydowanie się nie zgadzam (-2),
   (b) nie zgadzam się (-1),
   (c) zgadzam się (1),

(d) zdecydowanie się zgadzam (2).

Metryczka

1. (P) Płeć
   (a) kobieta (K),
   (b) mężczyzna (M).
2. (Wiek) Wiek
   (a) do 25 lat (1),
   (b) od 26 do 35 lat (2),
   (c) od 36 do 50 lat (3),
   (d) powyżej 50 lat (4).
3. (Wyk) Wykształcenie
   (a) zawodowe (1),
   (b) średnie (2),
   (c) wyższe (3).

2 Część 1

2.1 Zadanie 1

Na początku sporządźmy tabele liczności dla zmiennych $A1$ i $W1$. Zmienne te są ocenami zadowolenia z panującej atmosfery pracy i satysfakcji z otrzymywanego wynagrodzenia (zostały one uzyskane w pierwszym badanym okresie czasowym). Przedstawimy je zarówno uwzględniając wszystkie dane, jak i podzielimy je na różne podgrupy np. zadowolenie z atmosfery w pracy ze względu na wiek (rys. 1c) czy zadowolenie z wynagrodzenia ze względu na płeć, czyli tabela 2b.

```
> df_D_A1
# A tibble: 20 × 3
# Groups:   D [5]
   D     A1        n
   <chr> <chr> <int>
 1 D     A1        1
 2 O     -1        2
 3 O     0         4
 4 O     1        15
 5 O     2         5
 6 P     -1       10
 7 P     -2        9
 8 P     0        17
 9 P     1        51
10 P     2        11
11 S     -1        3
12 S     -2        3
13 S     0        14
14 S     1        15
15 S     2        10
16 Z     -1        2
17 Z     -2        2
18 Z     0         5
19 Z     1        19
20 Z     2         3
```

(a) Ze względu na dział

```
> df_P_A1
# A tibble: 11 × 3
# Groups:   P [3]
   P     A1        n
   <chr> <chr> <int>
 1 K     -1        7
 2 K     -2        3
 3 K     0        14
 4 K     1        36
 5 K     2        11
 6 M     -1       10
 7 M     -2       11
 8 M     0        26
 9 M     1        64
10 M     2        18
11 P     A1        1
```

(b) Ze względu na płeć

```
> df_Wiek_A1
# A tibble: 20 × 3
# Groups:   Wiek [5]
   Wiek  A1        n
   <chr> <chr> <int>
 1 1     -1        6
 2 1     -2        1
 3 1     0         3
 4 1     1        13
 5 1     2         3
 6 2     -1        7
 7 2     -2       11
 8 2     0        24
 9 2     1        50
10 2     2        12
11 3     -1        1
12 3     -2        2
13 3     0         5
14 3     1        25
15 3     2        12
16 4     -1        3
17 4     0         8
18 4     1        12
19 4     2         2
20 Wiek  A1        1
```

(c) Ze względu na wiek

```
> df_Wyk_A1
# A tibble: 16 × 3
# Groups:   Wyk [4]
   Wyk   A1        n
   <chr> <chr> <int>
 1 1     -1        6
 2 1     -2        5
 3 1     0         8
 4 1     1        19
 5 1     2         3
 6 2     -1       10
 7 2     -2        5
 8 2     0        26
 9 2     1        75
10 2     2        24
11 3     -1        1
12 3     -2        4
13 3     0         6
14 3     1         6
15 3     2         2
16 Wyk   A1        1
```

(d) Ze względu na wykształcenie

Rysunek 1: Tabele liczności dla zmiennej A1

```
> df_D_W1
# A tibble: 15 × 3
# Groups:   D [5]
   D     W1        n
   <chr> <chr> <int>
 1 D     W1        1
 2 O     -1        4
 3 O     -2        8
 4 O     1         1
 5 O     2        13
 6 P     -1       11
 7 P     -2       37
 8 P     1         1
 9 P     2        49
10 S     -1        2
11 S     -2       20
12 S     2        23
13 Z     -1        3
14 Z     -2        9
15 Z     2        19
```

(a) Ze względu na dział

```
> df_P_W1
# A tibble: 9 × 3
# Groups:   P [3]
  P     W1        n
  <chr> <chr> <int>
1 K     -1       10
2 K     -2       25
3 K     1         1
4 K     2        35
5 M     -1       10
6 M     -2       49
7 M     1         1
8 M     2        69
9 P     W1        1
```

(b) Ze względu na płeć

```
> df_Wiek_W1
# A tibble: 15 × 3
# Groups:   Wiek [5]
   Wiek  W1        n
   <chr> <chr> <int>
 1 1     -1        1
 2 1     -2        9
 3 1     2        16
 4 2     -1        9
 5 2     -2       42
 6 2     1         1
 7 2     2        52
 8 3     -1        6
 9 3     -2       12
10 3     2        27
11 4     -1        4
12 4     -2       11
13 4     1         1
14 4     2         9
15 Wiek  W1        1
```

(c) Ze względu na wiek

```
> df_Wyk_W1
# A tibble: 10 × 3
# Groups:   Wyk [4]
   Wyk   W1        n
   <chr> <chr> <int>
 1 1     -1        3
 2 1     -2       20
 3 1     2        18
 4 2     -1       17
 5 2     -2       45
 6 2     2        78
 7 3     -2        9
 8 3     1         2
 9 3     2         8
10 Wyk   W1        1
```

(d) Ze względu na wykształcenie

Rysunek 2: Tabele liczności dla zmiennej W1

2.2 Zadanie 2

Teraz przedstawmy tabele wielodzielcze dla następujących par zmiennych:

- zadowolenie z wynagrodzenia i płeć

```
> structable(P ~ W1, df)
   P  K  M  P
W1
-1   10 10  0
-2   25 49  0
1     1  1  0
2    35 69  0
W1    0  0  1
```

Rysunek 3: $W1$ i P

- zadowolenie z wynagrodzenia i stanowisko kierownicze

```
> structable(S ~ W1, df)
   S  0  1  S
W1
-1   18  2  0
-2   64 10  0
1     0  2  0
2    91 13  0
W1    0  0  1
```

Rysunek 4: $W1$ i S

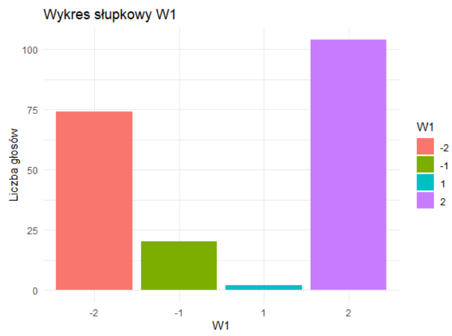
- ocena atmosfery w pracy i dział

```
> structable(D ~ A1, df)
   D  D  O  P  S  Z
A1
-1    0  2 10  3  2
-2    0  0  9  3  2
0     0  4 17 14  5
1     0 15 51 15 19
2     0  5 11 10  3
A1    1  0  0  0  0
```

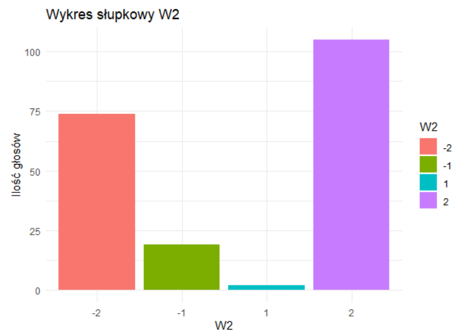
Rysunek 5: $A1$ i D

2.3 Zadanie 3

Sprawdźmy, jaki jest poziom zadowolenia pracowników z wynagrodzenia w obu badanych okresach ($W1$ i $W2$). W tym celu posłużymy się wykresem słupkowym rys. 6 i kołowym rys. 28.

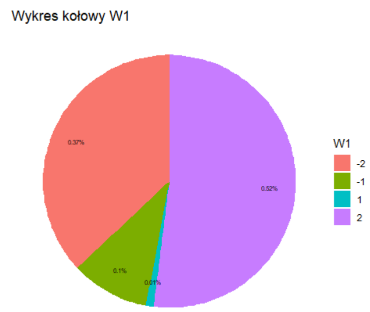


(a) W1

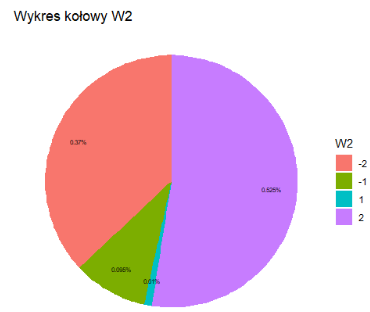


(b) W2

Rysunek 6: Wykres słupkowy poziomu zadowolenia z wynagrodzenia



(a) W1



(b) W2

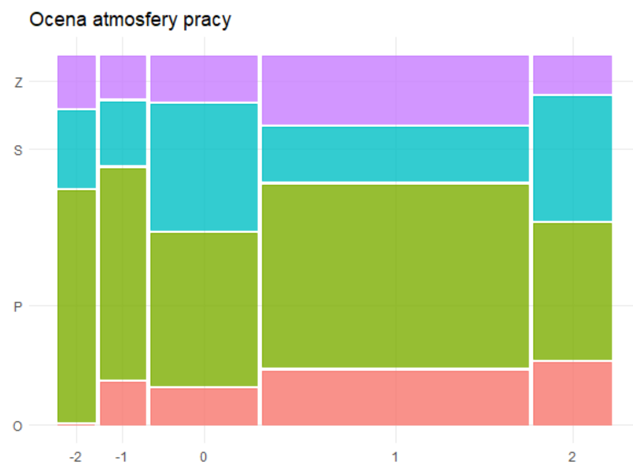
Rysunek 7: Wykres kołowy poziomu zadowolenia z wynagrodzenia

Jak możemy zauważyć, wyniki dla obu badanych okresów są bardzo do siebie zbliżone. Opinie dotyczące satysfakcji z wynagrodzenia są podzielone na dwa obozy - nieco więcej jest pracowników z pozytywną opinią. Zwróćmy także uwagę, że dominują odpowiedzi skrajne (tj. zdecydowanie się zgadzam, bądź zdecydowanie nie zgadzam).

2.4 Zadanie 4

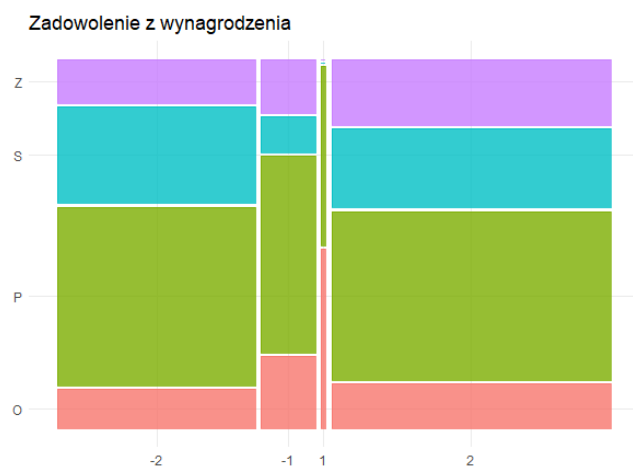
Korzystając z funkcji `mosaic` z biblioteki `vcd` sporządziliśmy wykresy odpowiadające parom zmiennych:

- dział i ocena atmosfery w pracy



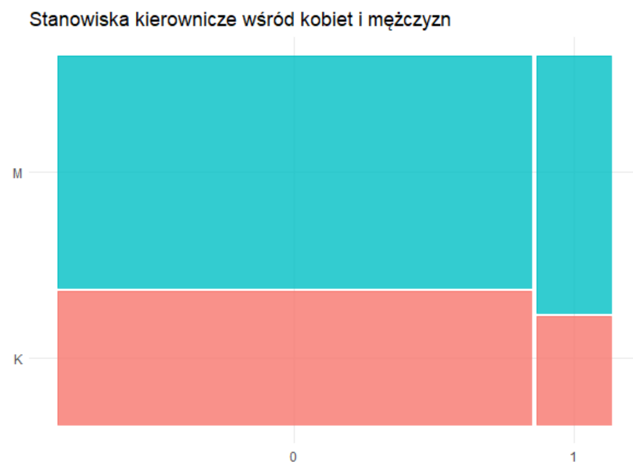
Rysunek 8: D i $A1$

- dział i zadowolenie z wynagrodzenia



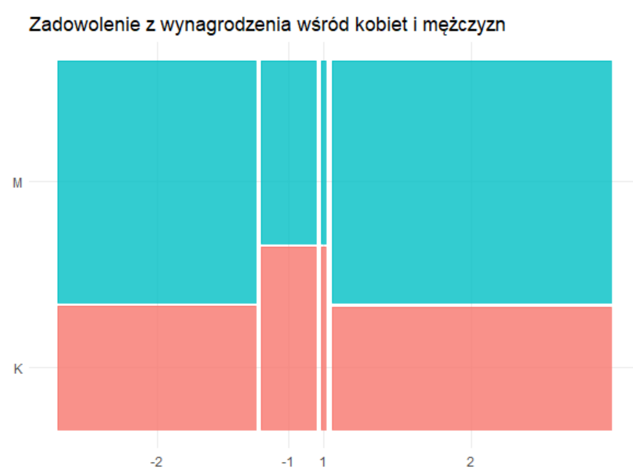
Rysunek 9: D i $W1$

- stanowisko kierownicze i płeć



Rysunek 10: S i P

- płeć i zadowolenie z wykształcenia



Rysunek 11: P i $W1$

3 Część 2

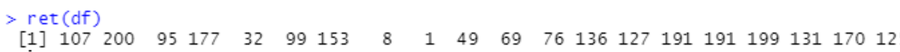
3.1 Zadanie 5

Następnie, używając funkcji `sample` z pakietu `stats`, napisaliśmy fragment programu, którego celem było wylosowanie próbki o rozmiarze około 10% liczby przypadków danej bazy danych.
Losowaliśmy w dwóch przypadkach:

- ze zwracaniem;

```r
library(stats)
  ret <- function(data){ #losowanie ze zwracaniem
    smp <- sample(nrow(data), size = 0.1*nrow(data), TRUE)
    return(smp)
  }
```

Powyższa funkcja zwraca nam przykładowy wyniki zaprezentowany na rys. 12, na którym widzimy, że zdarzyło się powtórzenie i dwa razy analizujemy osobę o tym samym ID.

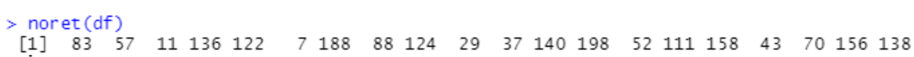


Rysunek 12: Losowanie ze zwracaniem

- bez zwracania;

```r
  noret <- function(data){ #losowanie bez zwracania
    smp <- sample(nrow(data), size = 0.1*nrow(data), FALSE)
    return(smp)
  }
```

Tym razem losujemy inną, zależną, próbkę. Nie mogą tutaj wystąpić powtórzenia.



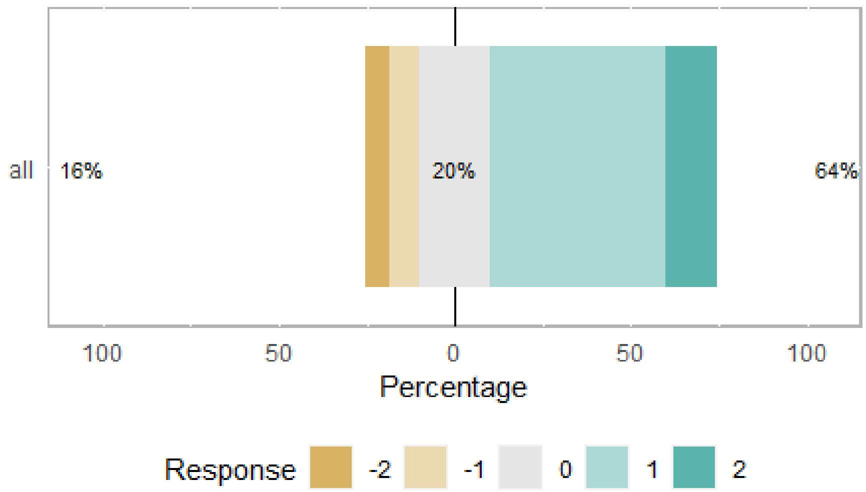
Rysunek 13: Losowanie bez zwracania

3.2 Zadanie 6

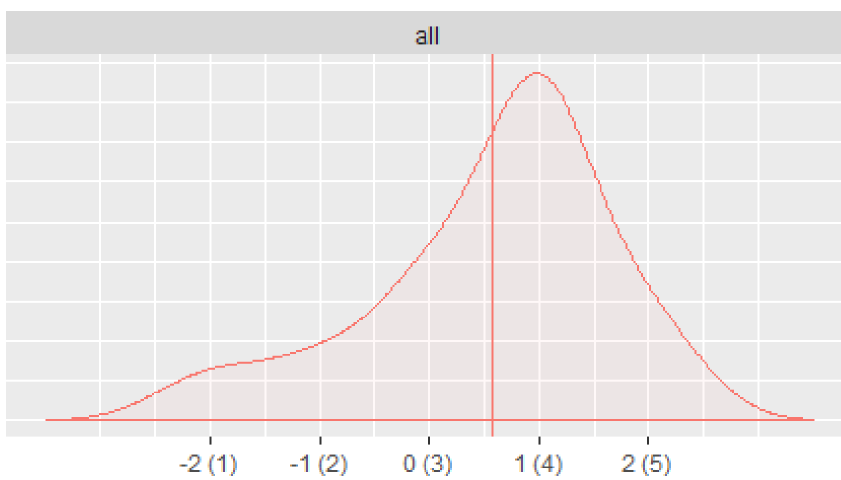
Następnie zilustrujemy dane dotyczące oceny atmosfery pracy. Ponieważ została w ankiecie zastosowana skala Likerta, możemy skorzystać do wizualizacji pakietu `likert`.

3.2.1 Pierwszy okres ankietyzacji - A1

- Cała grupa pracowników

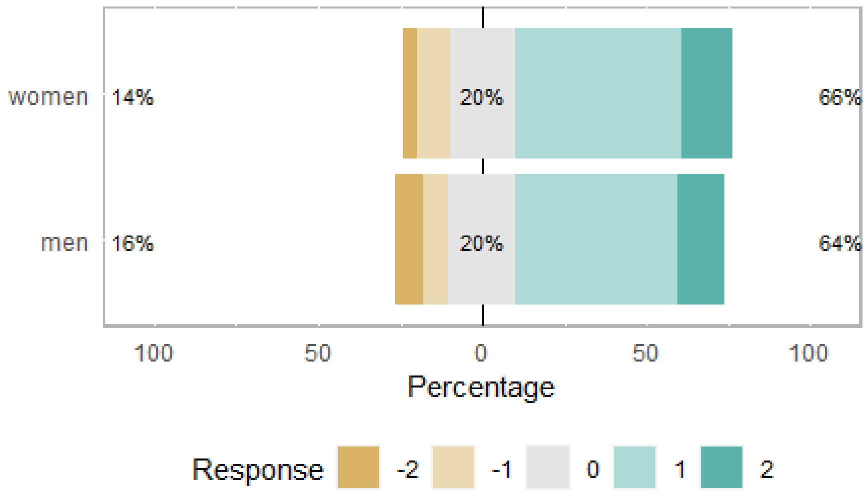


Rysunek 14: Wykres słupkowy dla całej grupy

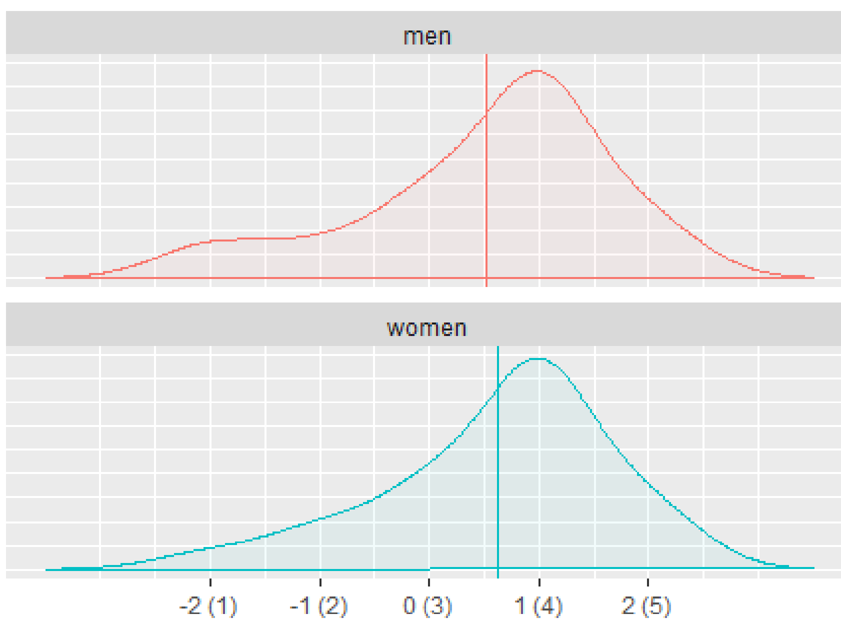


Rysunek 15: Gęstość ocen dla całej grupy

- Podział ze względu na płeć

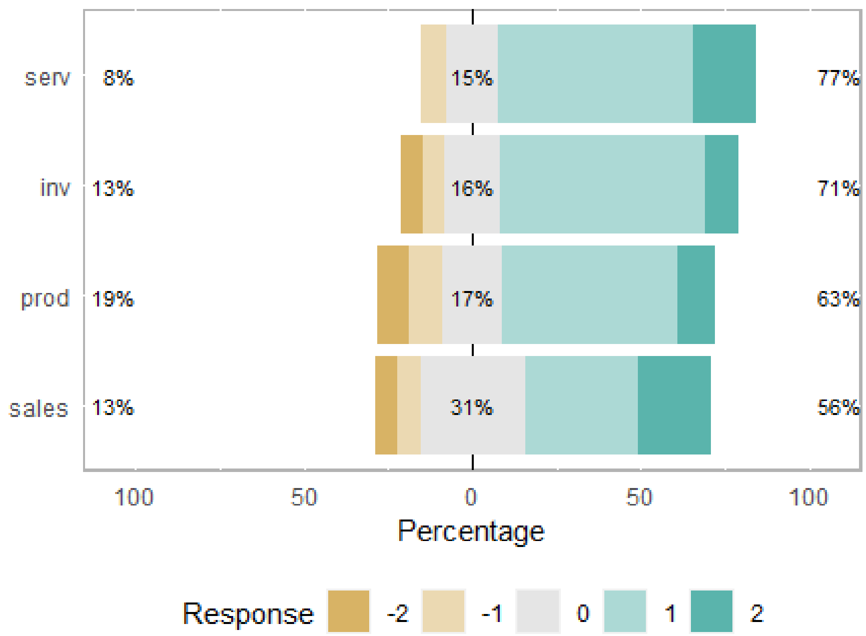


Rysunek 16: Wykres słupkowy dla podziału ze względu na płeć

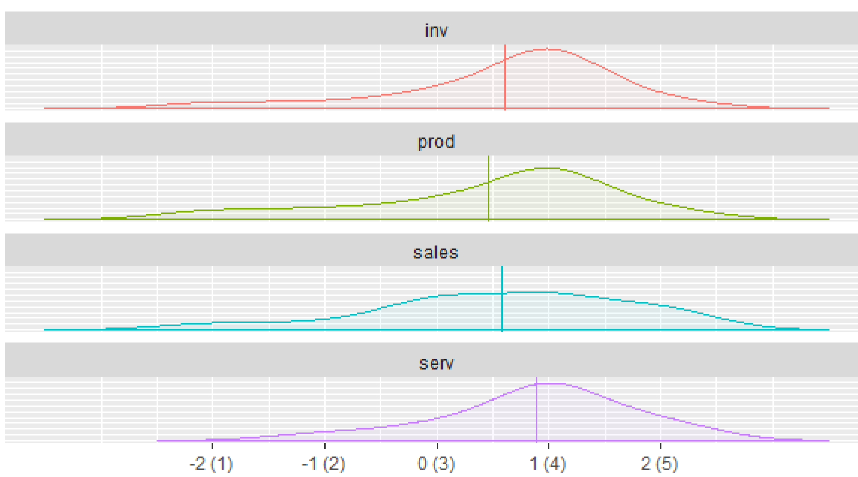


Rysunek 17: Wykres gęstości dla podziału ze względu na płeć

- Podział ze względu na dział



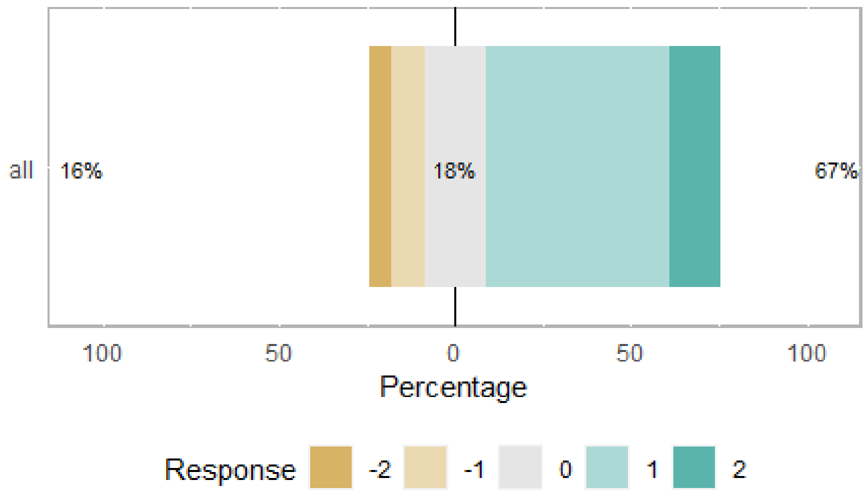
Rysunek 18: Wykres słupkowy dla podziału ze względu na dział



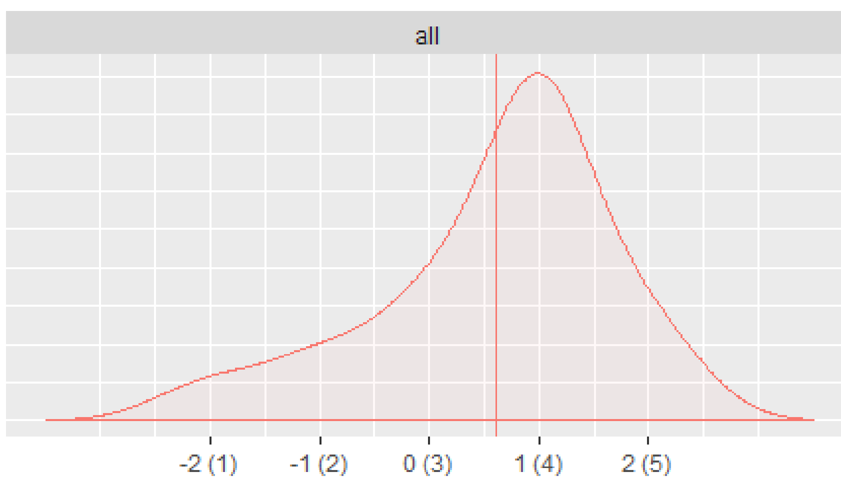
Rysunek 19: Wykres słupkowy dla podziału ze względu na dział

3.2.2 Drugi okres ankietyzacji - A2

- Cała grupa pracowników

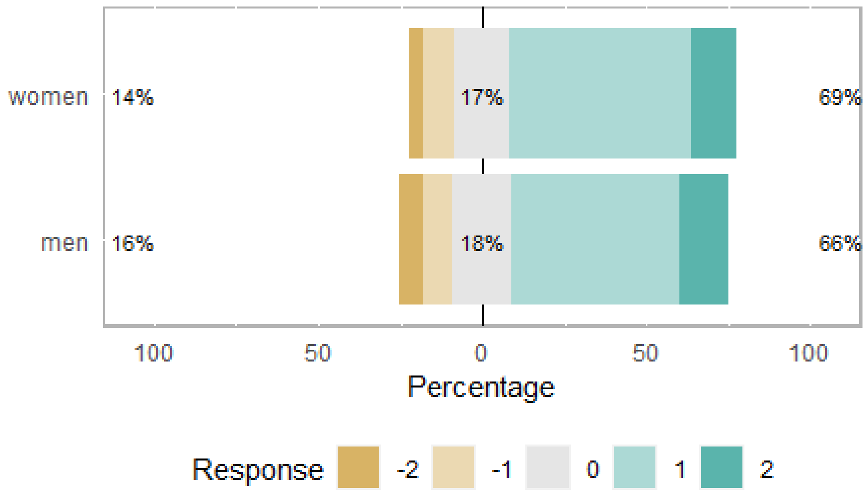


Rysunek 20: Wykres słupkowy dla całej grupy

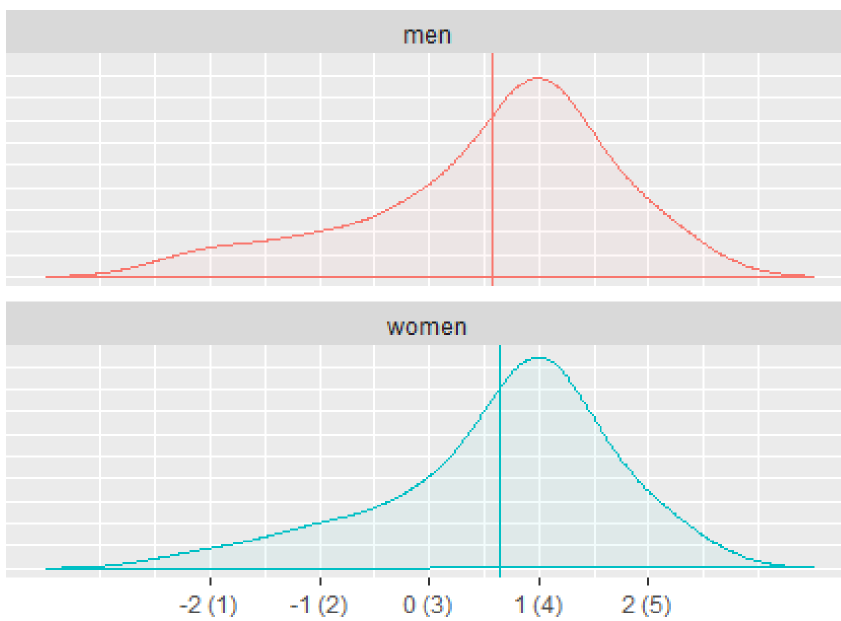


Rysunek 21: Gęstość ocen dla całej grupy

- Podział ze względu na płeć

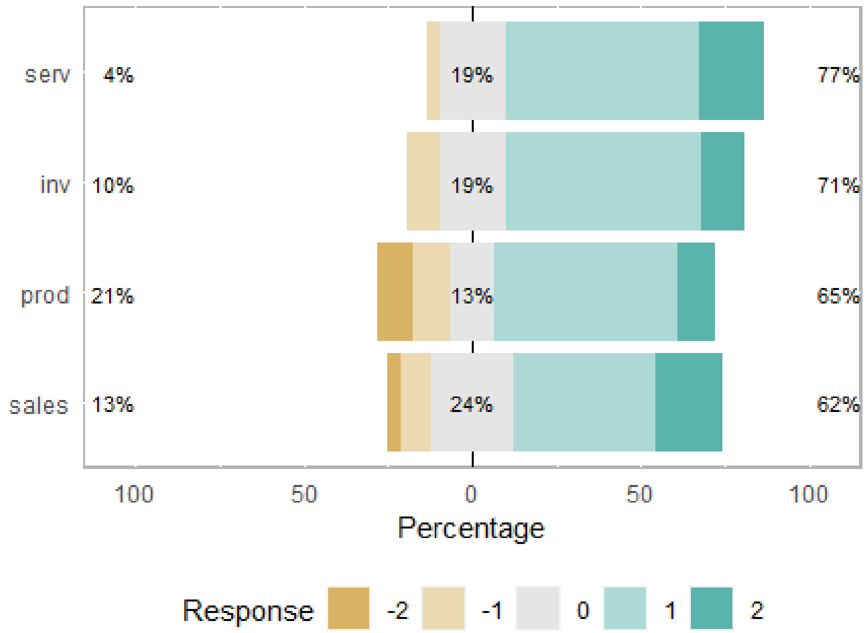


Rysunek 22: Wykres słupkowy dla podziału ze względu na płeć

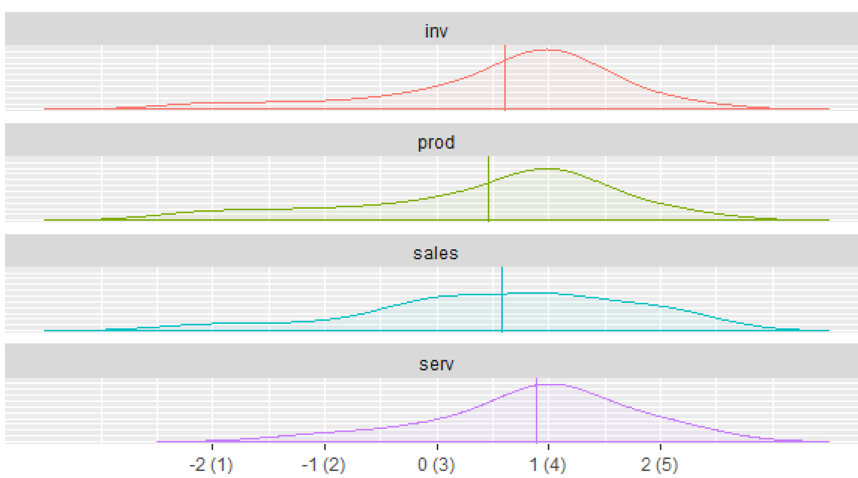


Rysunek 23: Wykres gęstości dla podziału ze względu na płeć

- Podział ze względu na dział



Rysunek 24: Wykres słupkowy dla podziału ze względu na dział



Rysunek 25: Wykres słupkowy dla podziału ze względu na dział

Jak możemy zauważyć (szczególnie to widać na wykresach słupkowych), opinie na temat atmosfery panującej w pracy nieznacznie się poprawiły. Zmalała ilość osób mających neutralną opinię, co spowodowało wzrost liczby pracowników z pozytywną opinią. Przykładowo w pierwszym okresie ankietyzacji 56% osób zatrudnionych w dziale sprzedaży miało pozytywną opinię na temat panującej atmosfery. W drugim okresie liczba ta wzrosła do 62%.

3.3 Zadanie 7

W tym zadaniu przyjrzymy się przedziałowi ufności Cloppera-Pearsona - dokładnej wersji przedziałów ufności dla zmiennych z rozkładu dwumianowego. Jest on zadany wzorem

$$\Big[p_{-\alpha}(X),\ p_{\alpha}(X)\Big],$$

gdzie

$$p_{-\alpha}(X) = \begin{cases} 0 \quad X = 0 \\ \text{kwantyl rzędu } {}^{\alpha}/_{2} \text{ rozkładu beta } \mathcal{B}e(X, n - X + 1) \end{cases}$$

$$p_{\alpha}(X) = \begin{cases} 1 \quad X = n \\ \text{kwantyl rzędu } 1 - {}^{\alpha}/_{2} \text{ rozkładu beta } \mathcal{B}e(X + 1, n - X) \end{cases}$$

Wyznaczymy za jego pomocą realizacje przedziałów ufności dla prawdopodobieństwa, że pracownik jest zadowolony ze swojego wynagrodzenia w pierwszym badanym okresie ($W1$). Uwzględnimy zarówno całą badaną grupę, jak i podgrupy pracowników ze względu na dział, oraz ze względu na stanowisko. W tym celu napiszemy funkcję, która wyznacza taki przedział ufności. Ponadto porównamy ją z funkcją wbudowaną `binom.confint` pakietu `binom`.

```
##Using built-in function
clopper1 <- function(alpha, poll){
  tbl <- table(poll)
  if((1 %in% poll) == FALSE){
    positive <- tbl[['2']]
  }
  else if((2 %in% poll) == FALSE){
    positive <- tbl[['1']]
  }
  else if(((1 %in% poll) == FALSE) &
          ((2 %in% poll) == FALSE)){
    positive <- 0
  }
  else{
    positive <- tbl[['1']] + tbl[['2']]
  }
  all <- sum(tbl)
  confint1 <- binom.confint(positive, all,
```

```
                          conf.level = 1 - alpha,
                          methods = 'exact')
  interval <- c(confint1$lower, confint1$upper)
  return(interval)
}

##Handmade version
clopper2 <- function(alpha, poll){
  tbl <- table(poll)
  if((1 %in% poll) == FALSE){
    positive <- tbl[['2']]
  }
  else if((2 %in% poll) == FALSE){
    positive <- tbl[['1']]
  }
  else if(((1 %in% poll) == FALSE) &
          ((2 %in% poll) == FALSE)){
    positive <- 0
  }
  else{
  positive <- tbl[['1']] + tbl[['2']]
  }
  all <- sum(tbl)
  left <- qbeta(alpha/2, positive, all - positive + 1)
  right <- qbeta(1 - alpha/2, positive + 1, all - positive)
  return(c(left, right))
}
```

Powyżej znajduje się fragment kodu napisanego w R zawierający dwie funkcje. Funkcja `clopper1` wyznacza przedziały ufności Cloppera-Pearsona wykorzystując funkcję wbudowaną z pakietu `binom`. Natomiast funkcja `clopper2` wykorzystuje wzór na krańce przedziałów i wylicza je "manualnie". W obu przypadkach zmienna `alpha` odpowiada za współczynnik istotności, a `poll` to zbiór opinii na temat wynagrodzenia z danej grupy pracowników.

```
> clopper1(alpha, df$W1)
[1] 0.4583305 0.6007671
> clopper2(alpha, df$W1)
[1] 0.4583305 0.6007671
```

Rysunek 26: Przykładowe wywołanie obu funkcji

Na powyższym obrazku znajduje się przykładowe wywołanie funkcji `clopper1` i `clopper2`. Jak się okazuje, zwracają one dokładnie takie same rezultaty. Poniżej przedstawiamy realizacje przedziałów ufności prawdopodobieństwa zadowolenia z wynagrodzenia dla poszczególnych grup pracowników. Za poziom istotności przedziałów przyjmujemy wartość 95%.

- Cała badana grupa: $[0.4583305,\ 0.6007671]$

- zaopatrzenia: $[0.4218696,\ 0.7815004]$
- dział produkcyjny: $[0.397266,\ 0.602734]$
- dział sprzedaży: $[0.3577404,\ 0.6629663]$
- dział obsługi kadrowo-płacowej: $[0.2992722,\ 0.7007278]$
- zwykli pracownicy: $[0.4488278,\ 0.6022889]$
- kierownicy: $[0.2866725,\ 0.6805035]$

Na podstawie powyższych wyników nie możemy jednoznacznie stwierdzić, w której grupie pracowników zadowolenie z wynagrodzenia jest najbardziej prawdopodobne. Możemy natomiast określić, który rezultat jest najdokładniejszy. Jak widać, rozstępy między krańcami przedziałów ufności mają różne długości. Im jest on węższy, tym lepiej oszacowane jest prawdopodobieństwo. Najlepsze wyniki uzyskaliśmy dla całej grupy i grupy zwykłych pracowników. Najmniej dokładne rezultaty są natomiast dla grupy kierowniczej oraz działu obsługi kadrowo-płacowej.

4 Część 3

4.1 Zadanie 8

Przedstawimy teraz propozycję algorytmu do generowania realizacji z rozkładu dwumianowego $\mathcal{B}(n, p)$. Następnie sprawdzimy jego poprawność porównując jego działanie z funkcją wbudowaną `rbinom`. Całość zwizualizujemy m.in. na histogramach.

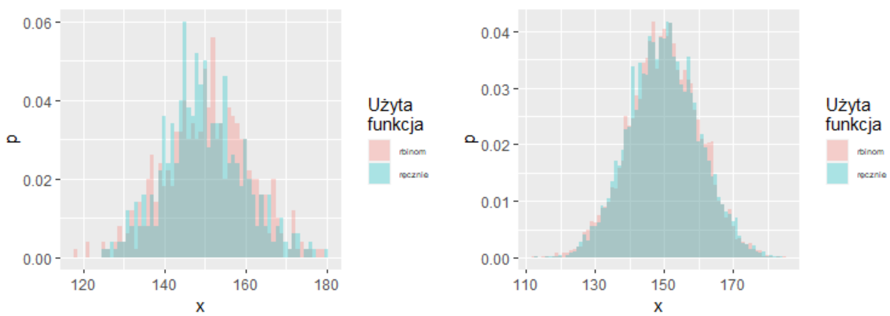
Propozycja algorytmu:

1. Generuj n zmiennych z rozkładu jednostajnego $\mathcal{U}(0, 1)$
2. Zlicz te zmienne, dla których zachodzi $X_i < p$. Oznacz jako Y_i
3. Powtórz kroki 1 i 2 N razy.

Powyższy algorytm generuje zmienne z rozkładu dwumianowego wykorzystując liczby pseudolosowe - imitację realizacji z rozkładu jednostajnego $\mathcal{U}(0, 1)$. Jego implementacja w R znajduje się poniżej.

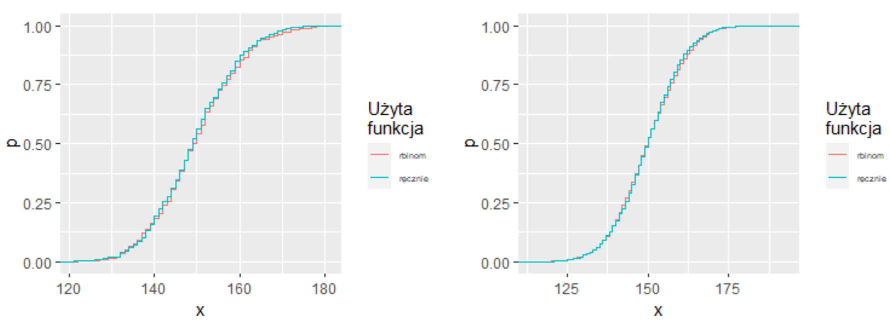
```r
## Binomial distribution sample generator
binomial <- function(n, p, N){
  bin <- c()
  for(i in 1:N){
    U <- runif(n, min = 0, max = 1)
    inds <- sum(U < p)
    bin <- append(bin, inds)
  }
  return(bin)
}
```

Sprawdźmy teraz wiarygodność naszego algorytmu. W tym celu wygenerujemy dwie próby zmiennych - jedną przy użyciu funkcji `rbinom`, a drugą wykorzystując naszą implementację. Następnie porównamy rozkłady otrzymanych prób na histogramach i dystrybuantach.



(a) $N = 500$ (b) $N = 5000$

Rysunek 27: Porównanie histogramów gęstości wygenerowanych prób z rozkładu dwumianowego $\mathcal{B}(500,\ 0.3)$



(a) $N = 500$ (b) $N = 5000$

Rysunek 28: Porównanie dystrybuant empirycznych wygenerowanych prób z rozkładu dwumianowego $\mathcal{B}(500,\ 0.3)$

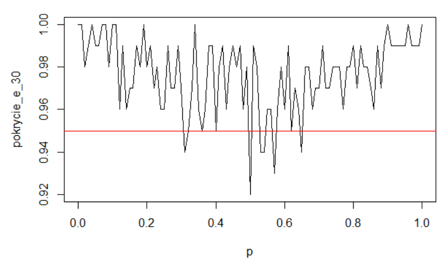
Na powyższych wykresach wyraźnie widać podobieństwo w rozkładach otrzymanych prób. Generalnie im większa jest liczba powtórzeń Monte Carlo N tym dokładniejszy jest napisany przez nas algorytm. Można zatem stwierdzić, że nasza implementacja jest dobrym zastępcą wbudowanej funkcji `rbinom`.

4.2 Zadanie 9

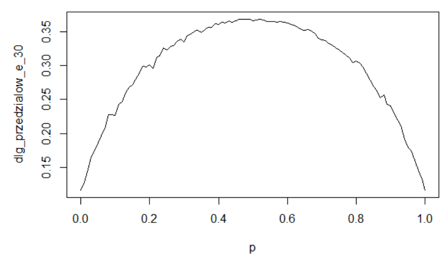
Przeprowadziliśmy symulację, której celem było porównanie prawdopodobieństwa pokrycia i długości przedziałów ufności Cloppera-Pearsona, Walda i

Agresti-Coulla. Uwzględniliśmy poziom ufności 0.95 oraz trzy rozmiary próby, kolejno 30, 100 oraz 1000. Wartości prawdopodobieństwa z zakresu $(0, 1)$. Zrobiliśmy to pisząc prostą funkcję a następnie wywołując ją dla konkretnych przypadków. Omówmy je po kolei.

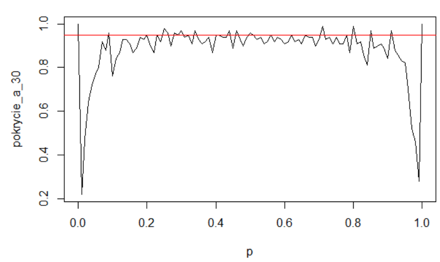
- Dla rozmiaru próby $n = 30$. Na rys. 29 możemy zobaczyć jak prezentują się prawdopodobieństwa pokrycia oraz długości przedziałów dla przedziałów ufności.



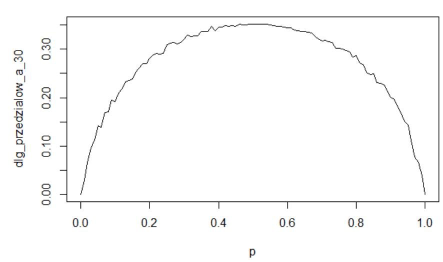
(a) Pokrycie dla Cloppera-Pearsona



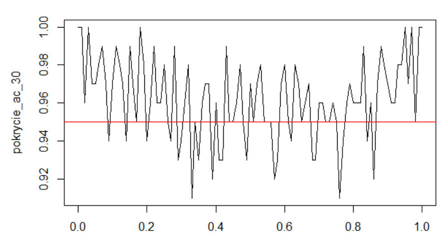
(b) Długości przedziałów dla Cloppera-Pearsona



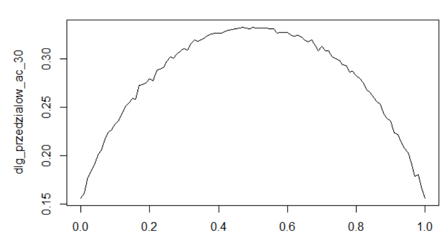
(c) Pokrycie dla Walda



(d) Długości przedziałów dla Walda



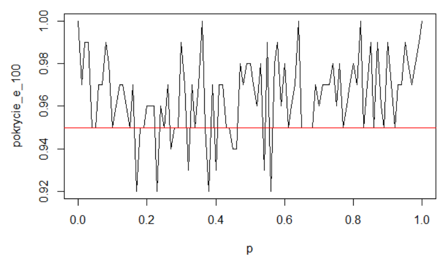
(e) Pokrycie dla Agresti-Coulla



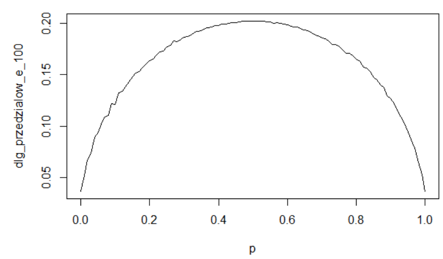
(f) Długości przedziałów dla Agresti-Coulla

Rysunek 29: Wyniki symulacji dla rozmiaru próby $n = 30$

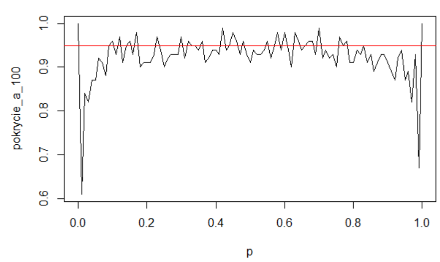
- Dla długości próby $n = 100$ wyniki mają się następująco:



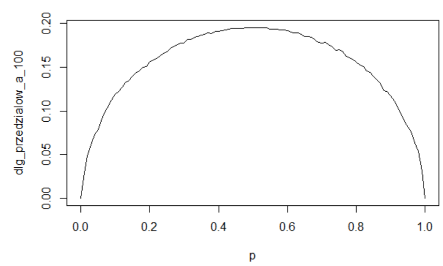
(a) Pokrycie dla Cloppera-Pearsona



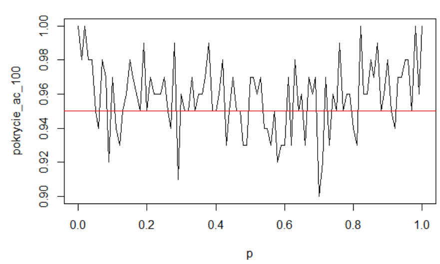
(b) Długości przedziałów dla Cloppera-Pearsona



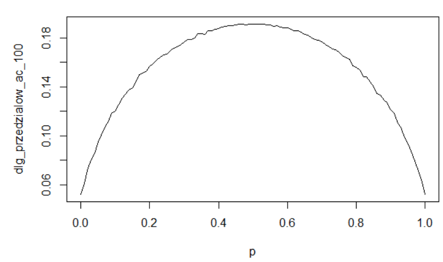
(c) Pokrycie dla Walda



(d) Długości przedziałów dla Walda



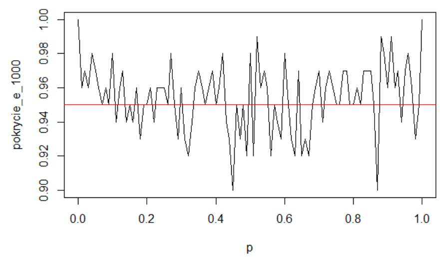
(e) Pokrycie dla Agresti-Coulla



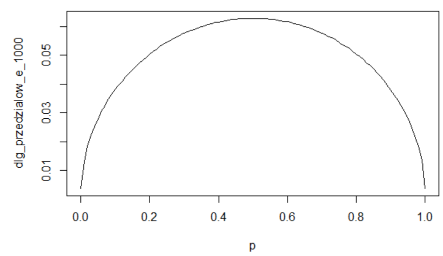
(f) Długości przedziałów dla Agresti-Coulla

Rysunek 30: Wyniki symulacji dla rozmiaru próby $n = 100$

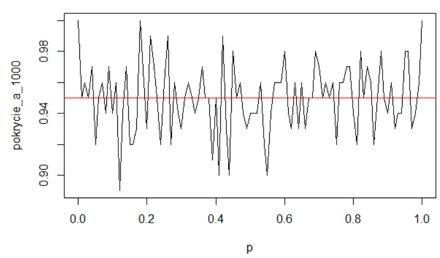
- Na sam koniec bierzemy pod uwagę $n = 1000$ a wyniki możemy podejrzeć na rys. 31.



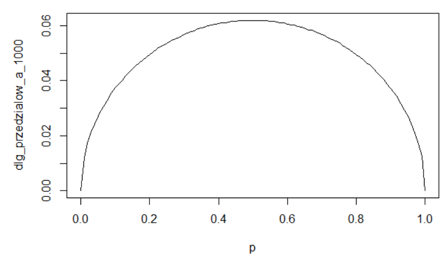
(a) Pokrycie dla Cloppera-Pearsona



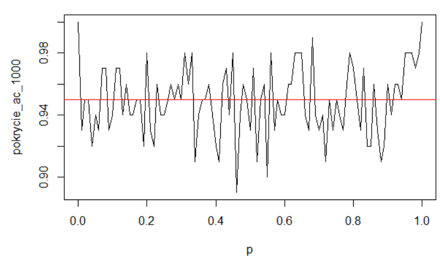
(b) Długości przedziałów dla Cloppera-Pearsona



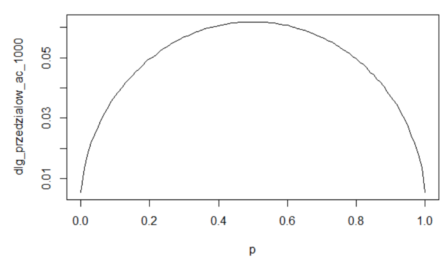
(c) Pokrycie dla Walda



(d) Długości przedziałów dla Walda



(e) Pokrycie dla Agresti-Coulla



(f) Długości przedziałów dla Agresti-Coulla

Rysunek 31: Wyniki symulacji dla rozmiaru próby $n = 1000$

Porównanie przedziałów ufności Cloppera-Pearsona, Walda i Agresti-Coulla może dostarczyć ważnych wniosków na temat skuteczności każdej z tych metod. W szczególności, porównanie prawdopodobieństwa pokrycia i długości przedziałów ufności dla różnych wartości prawdopodobieństwa z zakresu (0,1) może pomóc zrozumieć, jak zmieniają się wyniki w zależności od rozmiaru próby i poziomu ufności.
Przy poziomie ufności 0.95, można zauważyć, że przedziały ufności Cloppera-Pearsona i Agresti-Coulla mają zwykle wyższe prawdopodobieństwo pokrycia

niż przedział Walda. Jest to szczególnie widoczne dla małych wartości prawdopodobieństwa, gdzie przedziały ufności Cloppera-Pearsona i Agresti-Coulla mają prawdopodobieństwo pokrycia bliskie 1, podczas gdy przedział Walda może mieć znacznie niższe prawdopodobieństwo pokrycia.
Długość przedziałów ufności jest zwykle mniejsza dla przedziałów ufności Walda, co oznacza, że jest on bardziej precyzyjny. Jednak, dla mniejszych wartości prawdopodobieństwa i mniejszych prób, długość przedziałów ufności Cloppera-Pearsona i Agresti-Coulla może być znacznie mniejsza, co oznacza, że te metody są bardziej precyzyjne.
Wnioski te mogą mieć ważne implikacje dla analizy danych, w szczególności w przypadku małych prób lub dużych wartości prawdopodobieństwa. W tych przypadkach, przedziały ufności Cloppera-Pearsona i Agresti-Coulla mogą być bardziej skuteczne w oszacowaniu prawdziwej wartości parametru, a przedział Walda może być bardziej skuteczny w oszacowaniu przedziału, w którym prawdziwa wartość parametru znajduje się z dużą pewnością.

5 Część 4

5.1 Zadania 10, 11

Ostatnim elementem naszego sprawozdania będzie wykorzystanie funkcji `binom.test` oraz `prop.test` do zweryfikowania kilku hipotez, które zaraz wymienimy. Wpierw omówmy jednak budowę obu tych funkcji.

- `binom.test` - testuje hipotezę zerową o prawdopodobieństwie sukcesu w próbach Bernoulliego. Przyjmuje następujące argumenty:
  - `x` - liczba sukcesów (bądź wektor długości 2 zawierający liczbę sukcesów i porażek).
  - `n` - długość próby, ignorujemy jeśli podajemy za parametr `x` wektor z liczbami sukcesów i porażek.
  - `p` - hipotetyczne prawdopodobieństwo sukcesu.
  - `alternative` - hipoteza alternatywna, może to być `"two-sided"`, `"greater"`, albo `"less"`.
  - `conf.level` - poziom istotności α, w naszym przypadku będzie równy 0.05.
- `prop.test` - testuje hipotezę zerową o równości parametrów proporcji w różnych próbach lub o tym, że parametry te osiągają pewne konkretne wartości. Parametry są następujące:
  - `x` - wektor zawierający liczbę sukcesów (bądź macierz dwukolumnowa zawierająca pary liczby sukcesów i porażek).
  - `n` - wektor długości prób.

- `p` - wektor prawdopodobieństw sukcesów. Musi mieć taką samą długość jak `x`.
- `alternative` - hipoteza alternatywna, może to być `"two-sided"`, `"greater"`, albo `"less"`. Używana wyłącznie gdy testujemy hipotezę o jednej proporcji, lub o równości dwóch parametrów proporcji.
- `conf.level` - poziom istotności α. Muszą być spełnione te same warunki jak w przypadku argumentu `alternative`.
- `correct` - wartość logiczna, mówi czy stosować poprawkę ciągłości Yatesa w przypadku gdy to możliwe.

Po przeanalizowaniu działania naszych funkcji zweryfikowaliśmy pięć, niżej wymienionych, hipotez na poziomie istotności 0.95.
Hipotezę zerową będziemy przyjmować gdy przeprowadzając test otrzymamy $p-value$ niemniejsze od naszego zadanego poziomu ufności $\alpha = 0.05$.

- Prawdopodobieństwo, że w korporacji pracuje kobieta, w znaczeniu - prawdopodobieństwo, że pracownik jest kobietą, wynosi 0.5;
  Zdecydowaliśmy się skorzystać z pierwszej funkcji ze względu na to, że mamy zadane prawdopodobieństwo i sprawdzamy, czy jest ono jemu równe. Jak można zauważyć na rys. 32, otrzymaliśmy bardzo małe $p-value$, trudno byłoby uzyskać dane jeszcze bardziej wspierające hipotezę alternatywną, czyli taką która mówi, że prawdopodobieństwo, że pracownik korporacji jest kobietą nie jest równe 0.5.

```
> binom.test(kobiety, wszyscy, p=0.5, alternative="two.sided", conf.level=0.95)

        Exact binomial test

data:  kobiety and wszyscy
number of successes = 71, number of trials = 200, p-value = 4.973e-05
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.2887838 0.4255862
sample estimates:
probability of success
                 0.355
```

Rysunek 32: Weryfikacja hipotezy

  Widzimy także, że prawdopodobieństwo, że pracownik jest kobietą wynosi zaledwie 0.355.

- Prawdopodobieństwo, że pracownik jest zadowolony ze swojego wynagrodzenia jest większe bądź równe 0.8;
  Postanowiliśmy skorzystać z tej samej funkcji i dostaliśmy jednoznaczny wynik. $P-value$ równe 1 dla hipotezy alternatywnej wskazuje na prawidłowość tej alternatywy. Widzimy też, rys. 33, że otrzymane prawdopodobieństwo równe jest 0.535, czyli zdecydowanie mniej niż 0.8.

```
> binom.test(zadowoleni, wszyscy, p=0.8, alternative="greater", conf.level=0.95)

        Exact binomial test

data:  zadowoleni and wszyscy
number of successes = 107, number of trials = 200, p-value = 1
alternative hypothesis: true probability of success is greater than 0.8
95 percent confidence interval:
 0.4743904 1.0000000
sample estimates:
probability of success
                 0.535
```

Rysunek 33: Weryfikacja hipotezy

- Prawdopodobieństwo, że kobieta pracuje na stanowisku kierowniczym jest równe prawdopodobieństwu, że mężczyzna pracuje na stanowisku kierowniczym;
  Tym razem postanowiliśmy rozwiązać zagadnienie za pomocą funkcji `prop.test`, ze względu na to, że porównujemy ze sobą dwa prawdopodobieństwa nie znając ich wartości. Ustawiając argument `correct` jako `FALSE` nie korzystamy z korekcji ciągłości.

```
> prop.test(c(kierowniczki, kierownicy), c(wszyscy, wszyscy), correct = FALSE, alternative
 = "two.sided")

        2-sample test for equality of proportions without continuity correction

data:  c(kierowniczki, kierownicy) out of c(wszyscy, wszyscy)
X-squared = 4.8059, df = 1, p-value = 0.02836
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.103876448 -0.006123552
sample estimates:
prop 1 prop 2
 0.040  0.095
```

Rysunek 34: Weryfikacja hipotezy

  Na rys. 34 zauważamy, że $p - value$ jest mniejsze od poziomu ufności $\alpha = 0.05$ dlatego nie mamy podstaw do przyjęcia hipotezy zerowej i rozumiemy, że prawdopodobieństwo, że kobieta pracuje na stanowisku kierowniczym nie jest równe prawdopodobieństwu, że mężczyzna pracuje na stanowisku kierowniczym.

- Prawdopodobieństwo, że kobieta jest zadowolona ze swojego wynagrodzenia jest równe prawdopodobieństwu, że mężczyzna jest zadowolony ze swojego wynagrodzenia;
  Podglądając wynik kolejnego testu (rys. 35) ponownie odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej: prawdopodobieństwo, że kobieta jest zadowolona ze swojego wynagrodzenia jest różne od prawdopodobieństwa, że mężczyzna jest zadowolony ze swojego wynagrodzenia ze względu na to, że $p - value$ nie jest wystarczająco wysokie.

```
> prop.test(c(zad_kobiety, zad_mezczyzni), c(wszyscy, wszyscy), correct = FALSE, alternati
ve = "two.sided")

	2-sample test for equality of proportions without continuity correction

data:  c(zad_kobiety, zad_mezczyzni) out of c(wszyscy, wszyscy)
X-squared = 13.894, df = 1, p-value = 0.0001934
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.25023872 -0.07976128
sample estimates:
prop 1 prop 2
 0.185  0.350
```

Rysunek 35: Weryfikacja hipotezy

- Prawdopodobieństwo, że kobieta pracuje w dziale obsługi kadrowo-płacowej jest większe lub równe prawdopodobieństwu, że mężczyzna pracuje w dziale obsługi kadrowo-płacowej;
  Ponownie, korzystając z tej samej funkcji porównujemy dwa prawdopodobieństwa. Tak samo jak wcześniej, ze względu na wartość $p - value$ (rys. 36) nie mamy podstaw do przyjęcia hipotezy zerowej.

```
> prop.test(c(k_obsluga, m_obsluga), c(wszyscy, wszyscy), correct = FALSE, alternative =
 "greater")

	2-sample test for equality of proportions without continuity correction

data:  c(k_obsluga, m_obsluga) out of c(wszyscy, wszyscy)
X-squared = 16.454, df = 1, p-value = 2.492e-05
alternative hypothesis: greater
95 percent confidence interval:
 0.06029288 1.00000000
sample estimates:
prop 1 prop 2
 0.115  0.015
```

Rysunek 36: Weryfikacja hipotezy