

# Analiza danych ankietowych

## Raport 1

Klaudia Janicka 262268, Natalia Iwańska 262270

2023-04-13

## Część I

### Zadanie 1

Tablice licznosci dla zmiennej A1

```
#tablica licznosci dla A1
tabA1 <- personel %>% count(A1) %>% mutate(prop=n/sum(n))

#tablica licznosci dla A1 ze wzgledu na Wyk==1
tabA1wyk1 <- personel %>% filter(Wyk=="1") %>% count(A1) %>% mutate(prop=n/sum(n))
```

A1	n	prop
-2	14	0.070
-1	17	0.085
0	40	0.200
1	100	0.500
2	29	0.145

Tab. 1: Tablica licznosci dla A1.

A1	n	prop
-2	5	0.1219512
-1	6	0.1463415
0	8	0.1951220
1	19	0.4634146
2	3	0.0731707

Tab. 2: Tablica licznosci dla A1 ze wzgledu na Wyk=1.

A1	n	prop
-2	5	0.0357143
-1	10	0.0714286
0	26	0.1857143
1	75	0.5357143
2	24	0.1714286

Tab. 3: Tablica licznosci dla A1 ze wzgledu na Wyk=2.

A1	n	prop
-2	4	0.2105263
-1	1	0.0526316
0	6	0.3157895
1	6	0.3157895
2	2	0.1052632

Tab. 4: Tablica licznosci dla A1 ze wzgledu na Wyk=3.

A1	n	prop
-2	2	0.0645161
-1	2	0.0645161
0	5	0.1612903
1	19	0.6129032
2	3	0.0967742

Tab. 5: Tablica liczności dla A1 ze względu na D=Z.

A1	n	prop
-2	9	0.0918367
-1	10	0.1020408
0	17	0.1734694
1	51	0.5204082
2	11	0.1122449

Tab. 6: Tablica liczności dla A1 ze względu na D=P.

A1	n	prop
-2	3	0.0666667
-1	3	0.0666667
0	14	0.3111111
1	15	0.3333333
2	10	0.2222222

Tab. 7: Tablica liczności dla A1 ze względu na D=S.

A1	n	prop
-1	2	0.0769231
0	4	0.1538462
1	15	0.5769231
2	5	0.1923077

Tab. 8: Tablica liczności dla A1 ze względu na D=O.

A1	n	prop
-2	3	0.0422535
-1	7	0.0985915
0	14	0.1971831
1	36	0.5070423
2	11	0.1549296

Tab. 9: Tablica liczności dla A1 ze względu na P=k.

A1	n	prop
-2	11	0.0852713
-1	10	0.0775194
0	26	0.2015504
1	64	0.4961240
2	18	0.1395349

Tab. 10: Tablica liczności dla A1 ze względu na P=m.

### Tablice liczności dla zmiennej W1

A1	n	prop
-2	14	0.070
-1	17	0.085
0	40	0.200
1	100	0.500
2	29	0.145

Tab. 11: Tablica liczności dla W1.

A1	n	prop
-2	5	0.1219512
-1	6	0.1463415
0	8	0.1951220
1	19	0.4634146
2	3	0.0731707

Tab. 12: Tablica liczności dla W1 ze względu na Wyk=1.

A1	n	prop
-2	5	0.0357143
-1	10	0.0714286
0	26	0.1857143
1	75	0.5357143
2	24	0.1714286

Tab. 13: Tablica liczności dla W1 ze względu na Wyk=2.

A1	n	prop
-2	4	0.2105263
-1	1	0.0526316
0	6	0.3157895
1	6	0.3157895
2	2	0.1052632

Tab. 14: Tablica liczności dla W1 ze względu na Wyk=3.

A1	n	prop
-2	2	0.0645161
-1	2	0.0645161
0	5	0.1612903
1	19	0.6129032
2	3	0.0967742

Tab. 15: Tablica liczności dla W1 ze względu na D=Z.

A1	n	prop
-2	9	0.0918367
-1	10	0.1020408
0	17	0.1734694
1	51	0.5204082
2	11	0.1122449

Tab. 16: Tablica liczności dla W1 ze względu na D=P.

A1	n	prop
-2	3	0.0666667
-1	3	0.0666667
0	14	0.3111111
1	15	0.3333333
2	10	0.2222222

Tab. 17: Tablica liczności dla W1 ze względu na D=S.

A1	n	prop
-1	2	0.0769231
0	4	0.1538462
1	15	0.5769231
2	5	0.1923077

Tab. 18: Tablica liczności dla W1 ze względu na D=O.

A1	n	prop
-2	3	0.0422535
-1	7	0.0985915
0	14	0.1971831
1	36	0.5070423
2	11	0.1549296

Tab. 19: Tablica liczności dla W1 ze względu na P=k.

A1	n	prop
-2	11	0.0852713
-1	10	0.0775194
0	26	0.2015504
1	64	0.4961240
2	18	0.1395349

Tab. 20: Tablica liczności dla W1 ze względu na P=m.

## Zadanie 2

### Tabele wielozmiennych

```
kable(structable(W1~P, personel) %>% addmargins(),
      caption="Tabela wielozmiennych uwzględniająca zmienną W1 i P." %>%
      column_spec(1, border_left = TRUE) %>%
```

```
column_spec(6, border_right = TRUE) %>%
kable_styling(latex_options = "HOLD_position")
```

Tab. 21: Tabela wielodzielcza uwzględniająca zmienną W1 i P.

	-2	-1	1	2	Sum
K	25	10	1	35	71
M	49	10	1	69	129
Sum	74	20	2	104	200

Tab. 22: Tabela wielodzielcza uwzględniająca zmienną W1 i S.

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

Tab. 23: Tabela wielodzielcza uwzględniająca zmienną A1 i D.

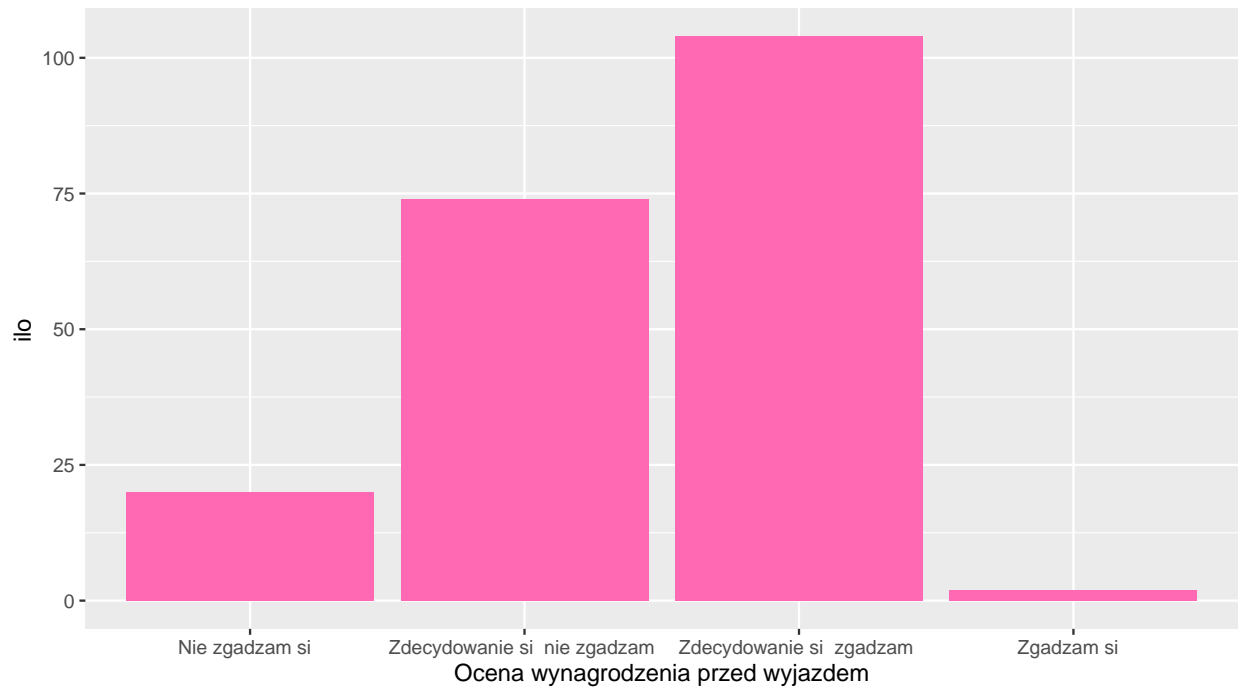
	-2	-1	0	1	2	Sum
O	0	2	4	15	5	26
P	9	10	17	51	11	98
S	3	3	14	15	10	45
Z	2	2	5	19	3	31
Sum	14	17	40	100	29	200

## Zadanie 3

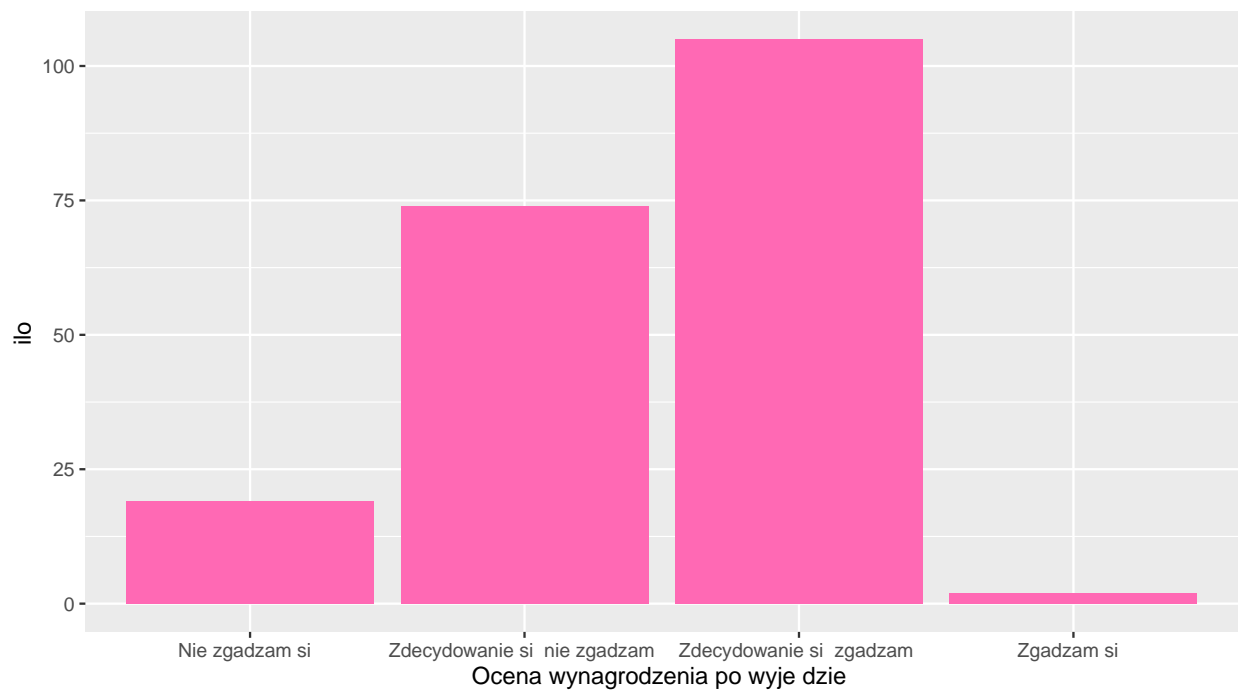
### Wykres słupkowy

```
daneW1 <- personel %>% count(W1) %>% data.frame()

ggplot(daneW1, aes(x=W1, y=n)) +
  geom_bar(stat = "identity", fill="hotpink") +
  xlab('Ocena wynagrodzenia przed wyjazdem') +
  ylab('ilość')
```



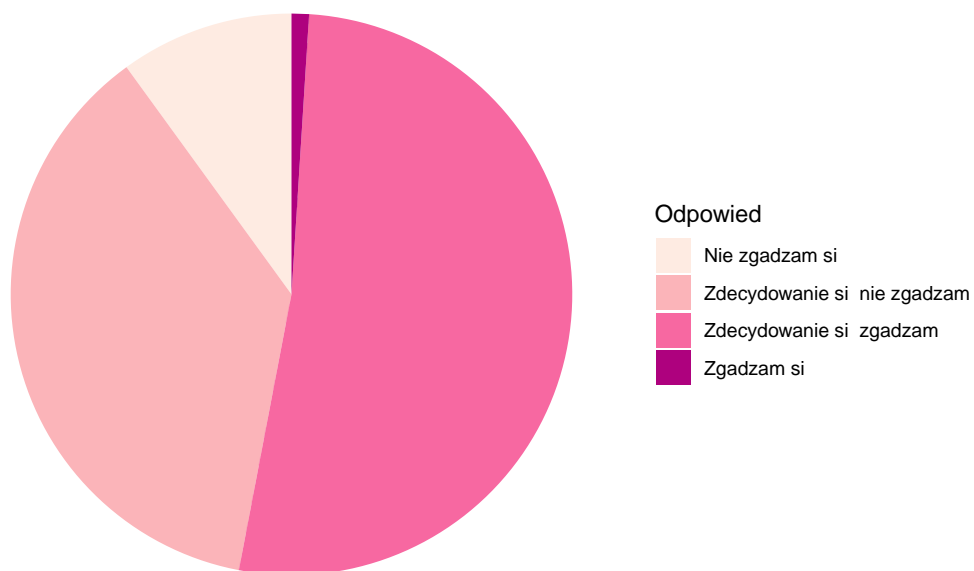
Wykres 1: Wykres słupkowy dla oceny wynagrodzenia przed wyjazdem



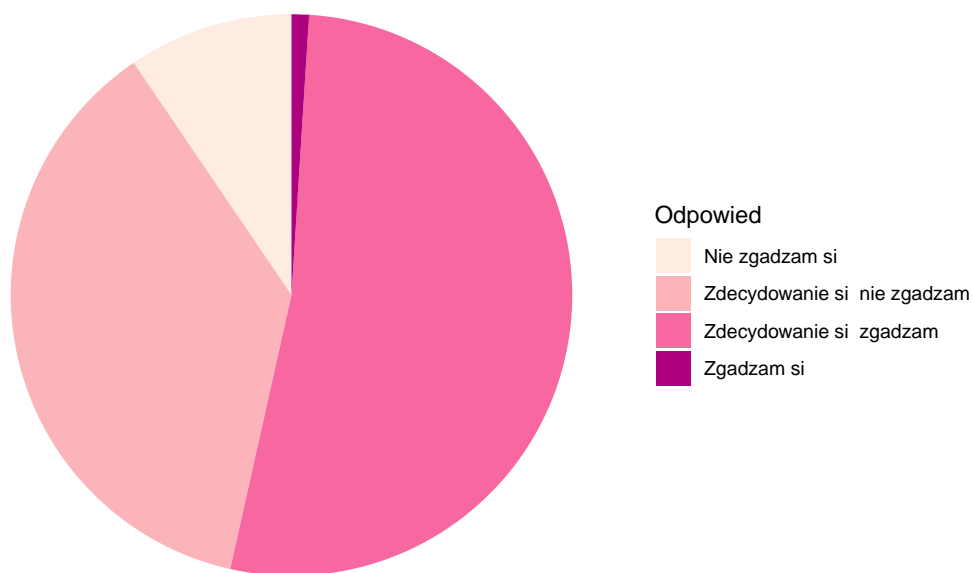
Wykres 2: Wykres słupkowy dla oceny wynagrodzenia po wyjeździe

## Wykres kołowy

```
ggplot(daneW1, aes(x="", y=n, fill=W1)) +  
  geom_bar(stat="identity", width=1) +  
  coord_polar("y", start=0) +  
  theme_void() +  
  scale_fill_brewer(palette="RdPu") +  
  guides(fill=guide_legend(title='Odpowiedź'))
```



Wykres 3: Wykres kołowy dla oceny wynagrodzenia przed wyjazdem



Wykres 4: Wykres kołowy dla oceny wynagrodzenia po wyjeździe

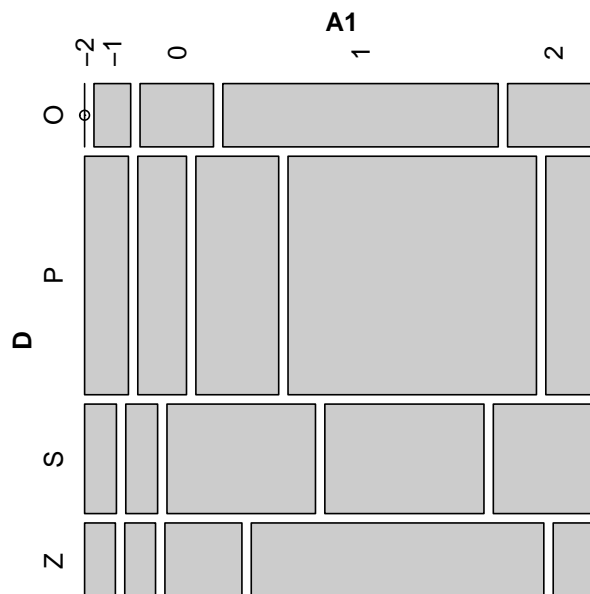
### Wnioski

Porównując oba wykresy kołowe (3, 4) i słupkowe (1, 2), nie widzimy znaczących zmian w odpowiedziach respondentów. Możemy zatem stwierdzić, że wyjazd nie wpłynął na zmianę oceny otrzymywanego wynagrodzenia. Ponadto można zauważyć, że większość pracowników jest co najmniej zadowolona ze swojego wynagrodzenia.

## Zadanie 4

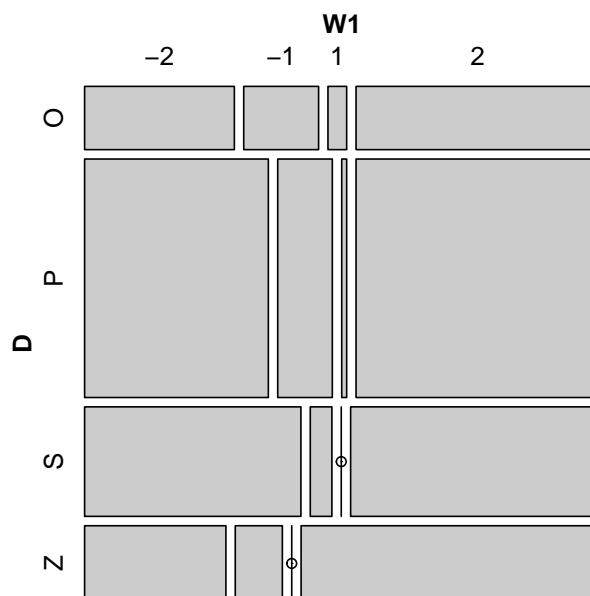
### Wykresy mozaikowe

```
mosaic(~D+A1, personel, labeling = vcd::labeling_border(rot_labels = c(90, 90)))
```



Wykres 5: Wykres mozaikowy dla działu oraz atmosfery.

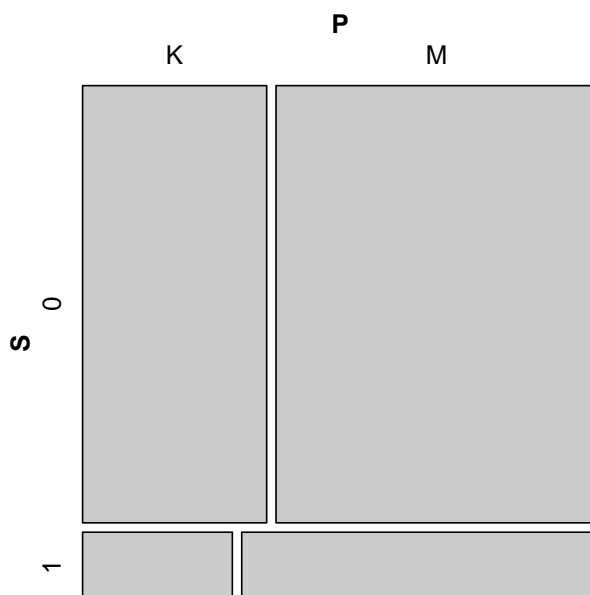
Na podstawie wykresu 5 możemy stwierdzić, że większość pracowników każdego z działów jest co najmniej zadowolona z atmosfery panującej w miejscu pracy. W dziale sprzedaży obserwujemy sporą, w stosunku do liczby pracowników, liczbę osób, które nie potrafią określić, czy są zadowolone, czy też nie.



Wykres 6: Wykres mozaikowy dla działu i zadowolenia z wynagrodzenia przed wyjazdem.

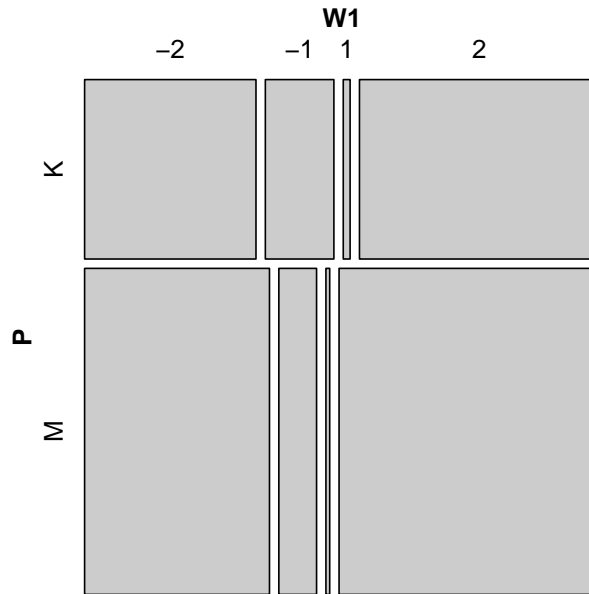


Jeśli chodzi o ocenę zadowolenia ze swojego wynagrodzenia (wykres 6), liczby są bardziej wyrównane. W dziale zaopatrzenia przeważają pozytywne odpowiedzi, w pozostałych trudniej zauważyć, które stanowią większość.



Wykres 7: Wykres mozikowy dla stanowiska kierowniczego i płci.

Zgodnie z wykresem 7 w firmie większość na stanowiskach kierowniczych stanowią mężczyźni. Jednak widać też, że w firmie pracuje ich więcej, zatem stosunek kierowniczek do kobiet i kierowników do mężczyzn może być podobny.



Wykres 8: Wykres mozaikowy dla płci i zadowolenia z wynagrodzenia przed wyjazdem.

Kobiety w firmie mają podzielone zdania na temat swojego wynagrodzenia, jednak jeśli już są zadowolone, to zdecydowanie. Natomiast mężczyźni są minimalnie bardziej zadowoleni niż nie, jednak liczby te również nie są od siebie mocno odległe (wykres 8).

## Część II

### Zadanie 5

Funkcja losująca ze zwracaniem i bez

```
f <- function(x='bez'){
  if (x=='zwracanie'){
    s <- sample(1:nrow(mtcars),3,replace=TRUE)
  } else{
    s <- sample(1:nrow(mtcars),3)
  }
  mtcars[s, ]
}
```

Tab. 24: Losowanie ze zwracaniem

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Duster 360	14.3	8	360	245	3.21	3.570	15.84	0	0	3	4
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1
Lincoln Continental	10.4	8	460	215	3.00	5.424	17.82	0	0	3	4

Tab. 25: Losowanie bez zwracania

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1

## Zadanie 6

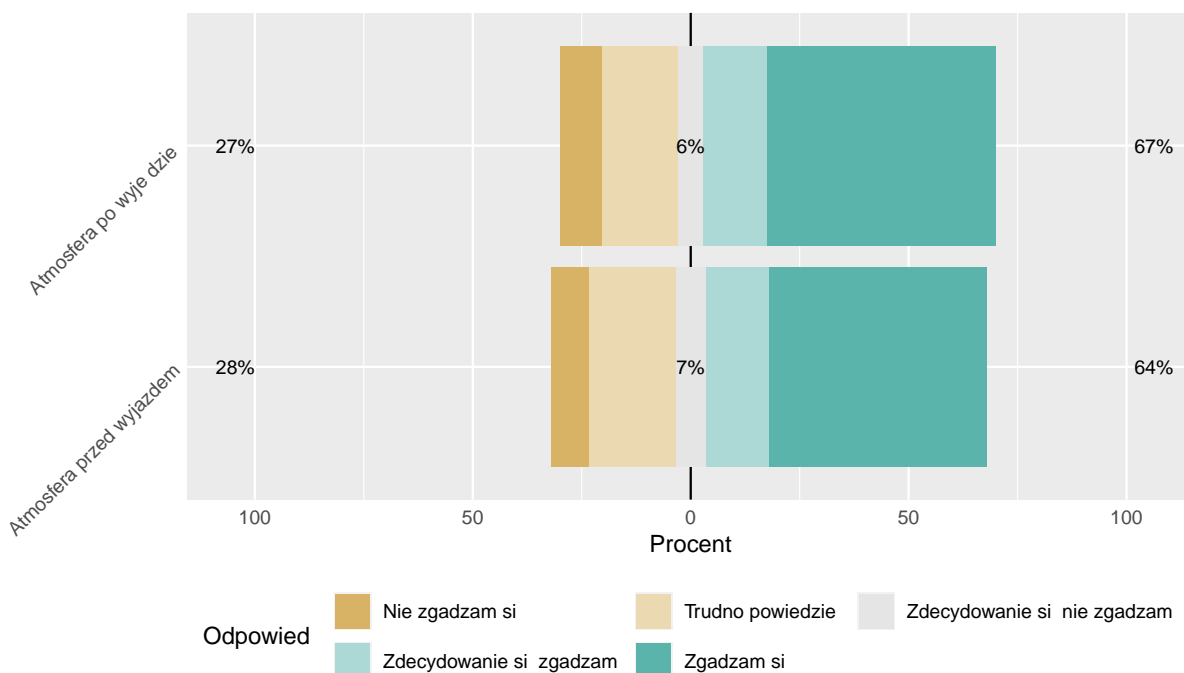
### Funkcja likert

```
df <- data.frame(personel$`Atmosfera przed wyjazdem`,personel$`Atmosfera po wyjeździe`)
colnames(df) <- c('Atmosfera przed wyjazdem', 'Atmosfera po wyjeździe')
likt_atmo <- likert(df)

summary(likt_atmo) %>% kable %>%
  column_spec(1, border_left = TRUE, border_right = FALSE) %>%
  column_spec(7, border_left = FALSE, border_right = TRUE) %>%
  kable_styling(latex_options = "HOLD_position")
```

	Item	low	neutral	high	mean	sd
2	Atmosfera po wyjeździe	27.0	6	67.0	3.830	1.456609
1	Atmosfera przed wyjazdem	28.5	7	64.5	3.775	1.447359

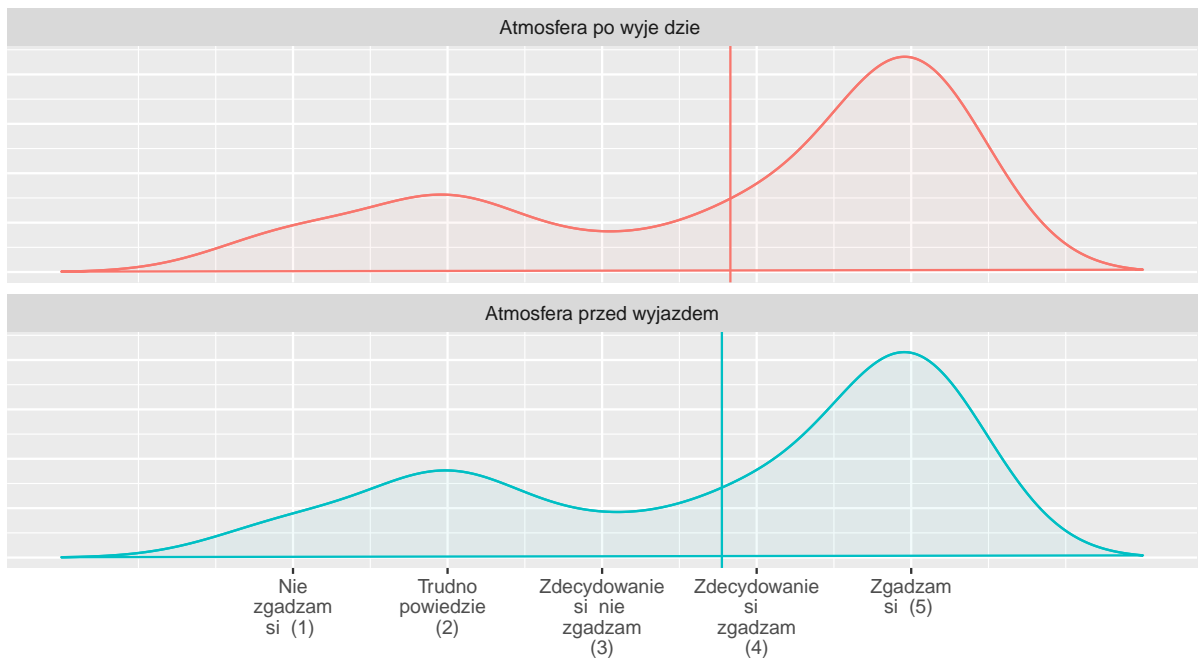
```
likert.bar.plot(likt_atmo) +
  guides(fill=guide_legend(nrow=2,byrow=TRUE,title='Odpowiedź')) +
  theme(axis.text.y=element_text(angle=45,hjust=1)) +
  ylab("Procent")
```



```
plot(likert_atmo,type='heat') +
  theme(axis.text.x=element_text(angle=90,hjust=1),axis.text.y=element_text(angle=45,hjust=1)) +
  guides(fill=guide_legend(title='Odpowiedź'))
```

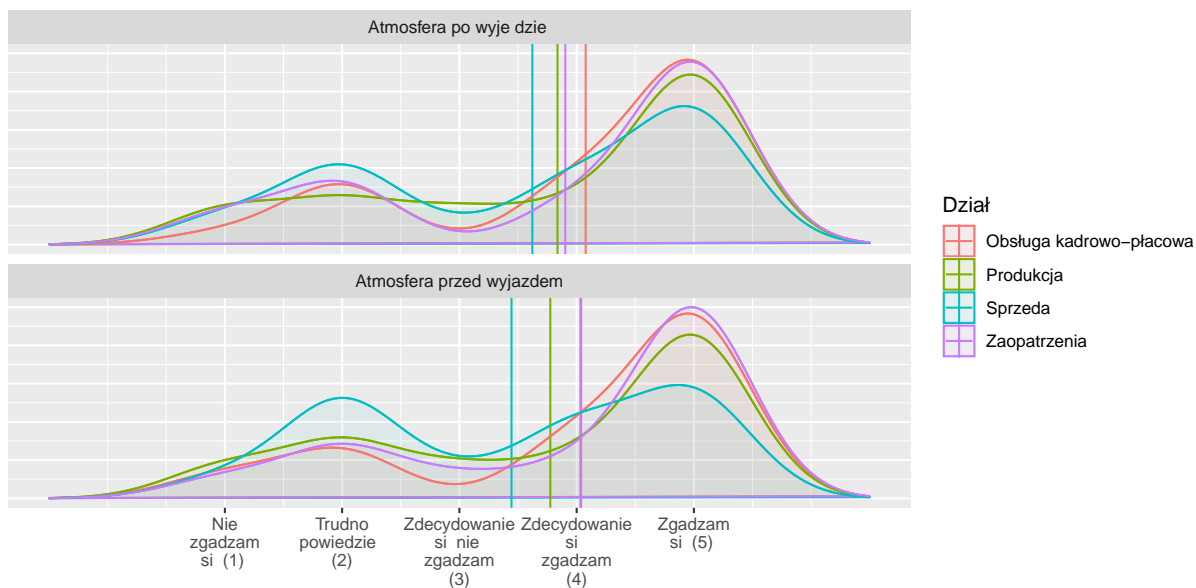
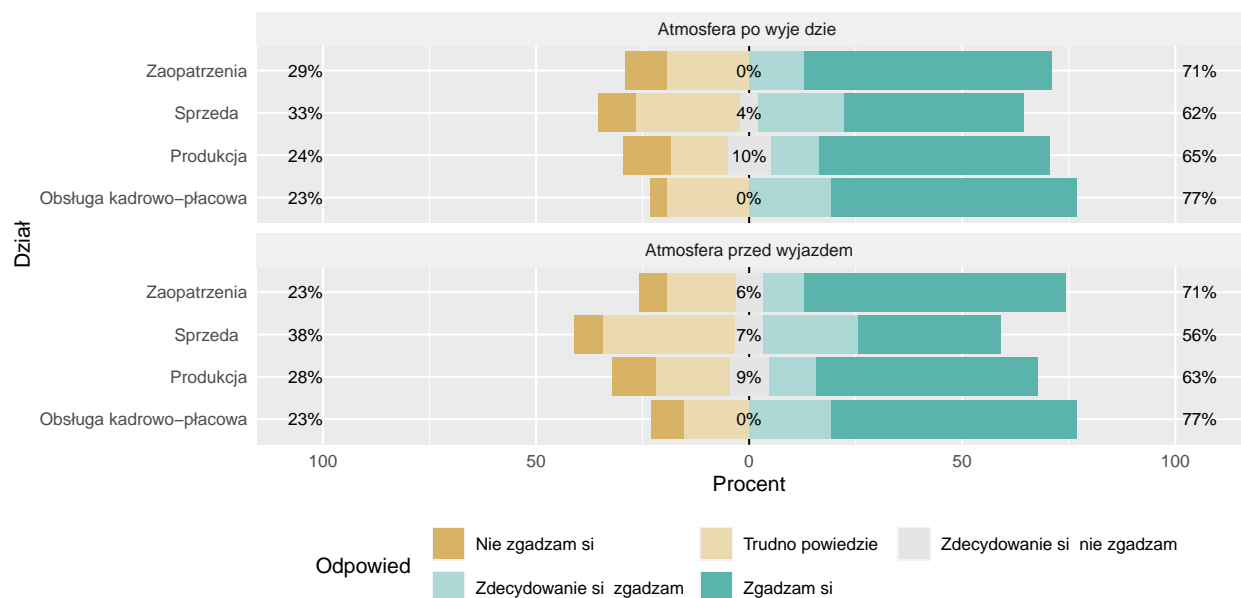


```
likert.density.plot(likert_atmo)
```



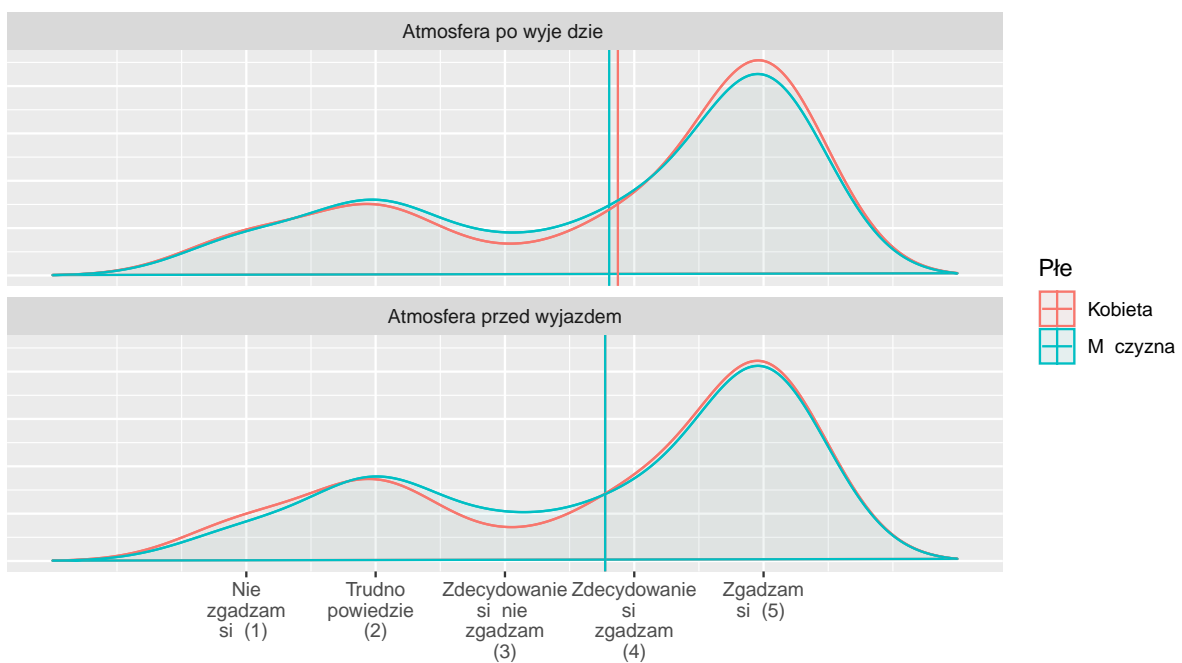
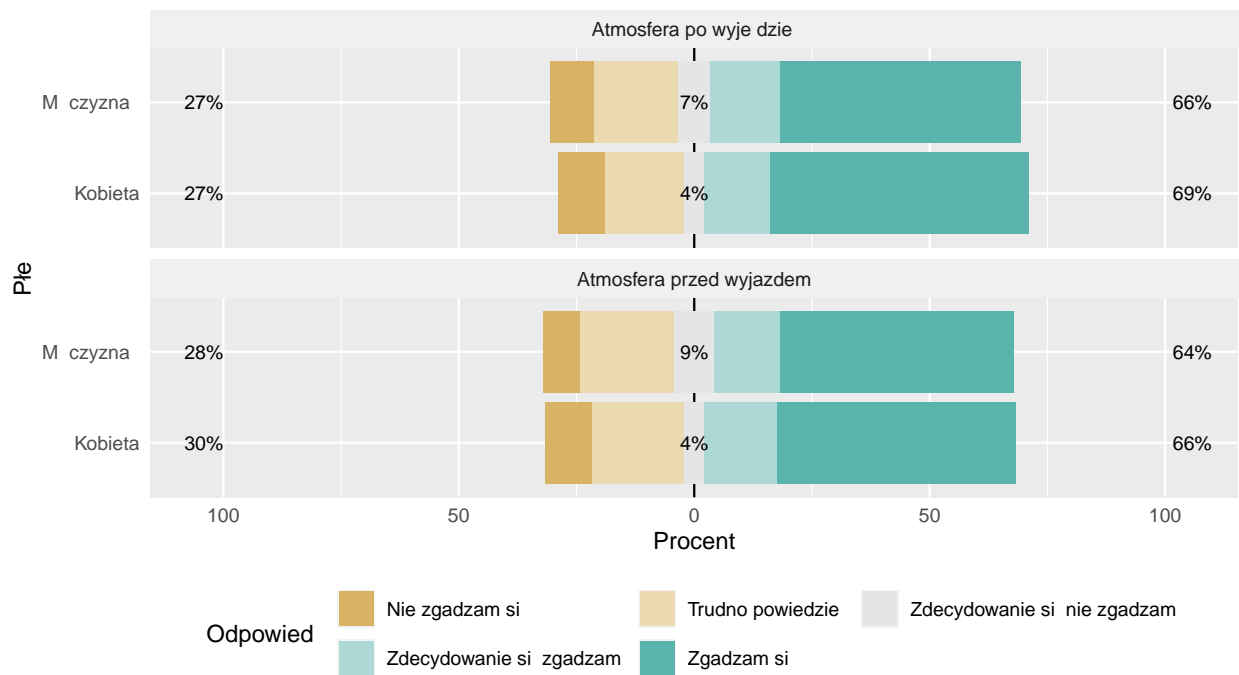
## Podgrupa ze względu na dział

Group	Item	low	neutral	high	mean	sd
Obsługa kadrowo-płacowa	Atmosfera przed wyjazdem	23.07692	0.000000	76.92308	4.038462	1.399450
Obsługa kadrowo-płacowa	Atmosfera po wyjeździe	23.07692	0.000000	76.92308	4.076923	1.324329
Produkcja	Atmosfera przed wyjazdem	27.55102	9.183673	63.26531	3.775510	1.482062
Produkcja	Atmosfera po wyjeździe	24.48980	10.204082	65.30612	3.836735	1.476230
Sprzedaż	Atmosfera przed wyjazdem	37.77778	6.666667	55.55556	3.444444	1.407053
Sprzedaż	Atmosfera po wyjeździe	33.33333	4.444444	62.22222	3.622222	1.466116
Zaopatrzenia	Atmosfera przed wyjazdem	22.58065	6.451613	70.96774	4.032258	1.401995
Zaopatrzenia	Atmosfera po wyjeździe	29.03226	0.000000	70.96774	3.903226	1.513381



## Podgrupa ze względu na płeć

Group	Item	low	neutral	high	mean	sd
Kobieta	Atmosfera przed wyjazdem	29.57746	4.225352	66.19718	3.774648	1.485138
Kobieta	Atmosfera po wyjeździe	26.76056	4.225352	69.01408	3.873239	1.472894
Mężczyzna	Atmosfera przed wyjazdem	27.90698	8.527132	63.56589	3.775194	1.432002
Mężczyzna	Atmosfera po wyjeździe	27.13178	6.976744	65.89147	3.806202	1.452786



## Zadanie 7

### Przedział ufności Cloppera-Pearsona

```
p.lower <- function(x, n, a){
  if(x == 0){
    return(0)
  }
  else{
    return(qbeta(a/2,x,n-x+1))
  }
}

p.upper <- function(x, n, a){
  if(x == n){
    return(1)
  }
  else{
    return(qbeta(1-a/2, x+1, n-x))
  }
}

clopper_pearson_ci <- function(x, n=NULL, a=0.05){
  if(is.null(n)){
    n <- length(x)
    x <- sum(x==1)
    return(data.frame(x=x, n=n, lower=p.lower(x, n, a), upper=p.upper(x, n, a)))
  }
  else{
    return(data.frame(x=x, n=n, lower=p.lower(x, n, a), upper=p.upper(x, n, a)))
  }
}
```

### Przykład użycia

- funkcja wbudowana

method	x	n	mean	lower	upper
exact	7	20	0.35	0.1539092	0.5921885

- funkcja `clopper_pearcon_ci`

x	n	lower	upper
7	20	0.1539092	0.5921885

### Zadowolenie z wynagrodzenia w całej badanej grupie

- funkcja `clopper_pearson_ci`

x	n	lower	upper
106	200	0.4583305	0.6007671

- funkcja wbudowana

method	x	n	mean	lower	upper
exact	106	200	0.53	0.4583305	0.6007671

### Zadowolenie z wynagrodzenia ze względu na dział

- funkcja `clopper_pearson_ci`

x	n	lower	upper	Dział
14	26	0.3337082	0.7341288	Obsługa kadrowo-płacowa
50	98	0.4071736	0.6126014	Produkcja
23	45	0.3577404	0.6629663	Sprzedaż
19	31	0.4218696	0.7815004	Zaopatrzenia

- funkcja wbudowana

method	x	n	mean	lower	upper
exact	14	26	0.5384615	0.3337082	0.7341288
exact	50	98	0.5102041	0.4071736	0.6126014
exact	23	45	0.5111111	0.3577404	0.6629663
exact	19	31	0.6129032	0.4218696	0.7815004

### Zadowolenie z wynagrodzenia ze względu na stanowisko

- funkcja `clopper_pearson_ci`

x	n	lower	upper	Stanowisko kierownicze
91	173	0.4488278	0.6022889	Nie
15	27	0.3532642	0.7452012	Tak

- funkcja wbudowana

method	x	n	mean	lower	upper
exact	91	173	0.5260116	0.4488278	0.6022889
exact	15	27	0.5555556	0.3532642	0.7452012

### Wnioski

TOTALNIE NIE MAM POJĘCIA CO TU NAPISAĆ!!!! :(

## Część III

### Zadanie 8

#### Generowanie rozkładu dwumianowego

Do wygenerowania liczb z rozkładu dwumianowego wykorzystamy poniższy algorytm:



1. Ustal  $n$  i  $p$ .
2. Generuj  $Y_i$  z rozkładu  $P(Y_i = 1) = p = 1 - P(Y_i = 0)$ .
3. Powtórz krok 2.  $n$  razy.
4. Wstaw  $X = \sum_{i=1}^n Y_i$ .
5. Powtórz kroki 2-4  $N$  razy.

Korzystamy z faktu, że suma zmiennych losowych z rozkładu Bernoulliego jest zmienną z rozkładu dwumianowego.

Aby to udowodnić, użyjemy funkcji charakterystycznych obu rozkładów.

Niech  $X$  będzie zmienną losową z rozkładu dwumianowego z parametrami  $n$  i  $p$ , a  $Y_i$  zmienną z rozkładu Bernoulliego ( $P(Y_i = 1) = p = 1 - P(Y_i = 0)$ ).

Funkcja charakterystyczna rozkładu Bernoulliego dana jest wzorem  $\phi_{Y_i}(t) = \mathbb{E}(e^{itY_i}) = 1 - p + pe^{it}$ .

Możemy zapisać funkcję charakterystyczną jako:

$$\phi_{\sum_i Y_i}(t) = \mathbb{E}(e^{it \sum_i Y_i}) = \mathbb{E}(e^{it(Y_1+Y_2+\dots+Y_n)}) = \mathbb{E}(e^{itY_1} e^{itY_2} \dots e^{itY_n}).$$

Korzystając z niezależności zmiennych:

$$\phi_{\sum_i Y_i}(t) = \mathbb{E}(e^{itY_1})\mathbb{E}(e^{itY_2})\dots\mathbb{E}(e^{itY_n}) = (\mathbb{E}e^{itY_1})^n = (1 - p + pe^{it})^n.$$

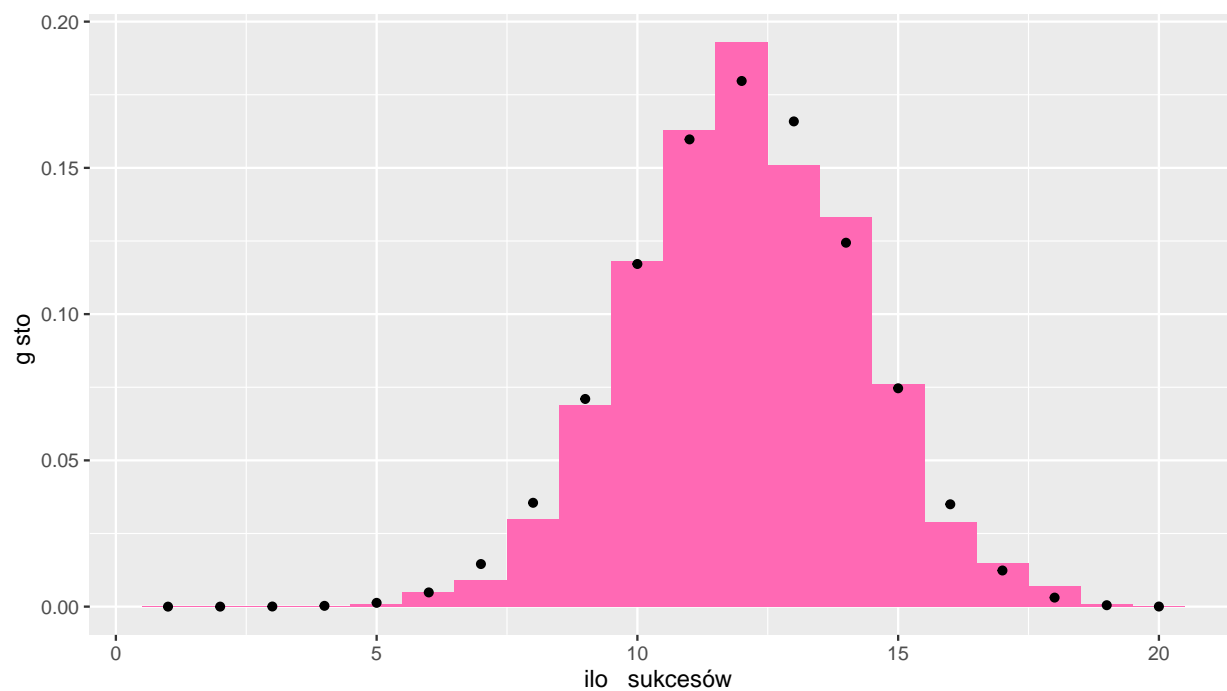
Otrzymamy wynik jest funkcją charakterystyczną rozkładu dwumianowego, co kończy dowód.

```
generuj_dwumianowy <- function(n, p, N){
  X <- rep(NA, N)
  for(i in 1:N){
    Y <- rep(NA, n)
    for(j in 1:n){
      Y[j] <- sample(c(0,1), size=1, prob=c(1-p,p))
    }
    X[i] <- sum(Y)
  }
  return(X)
}
```

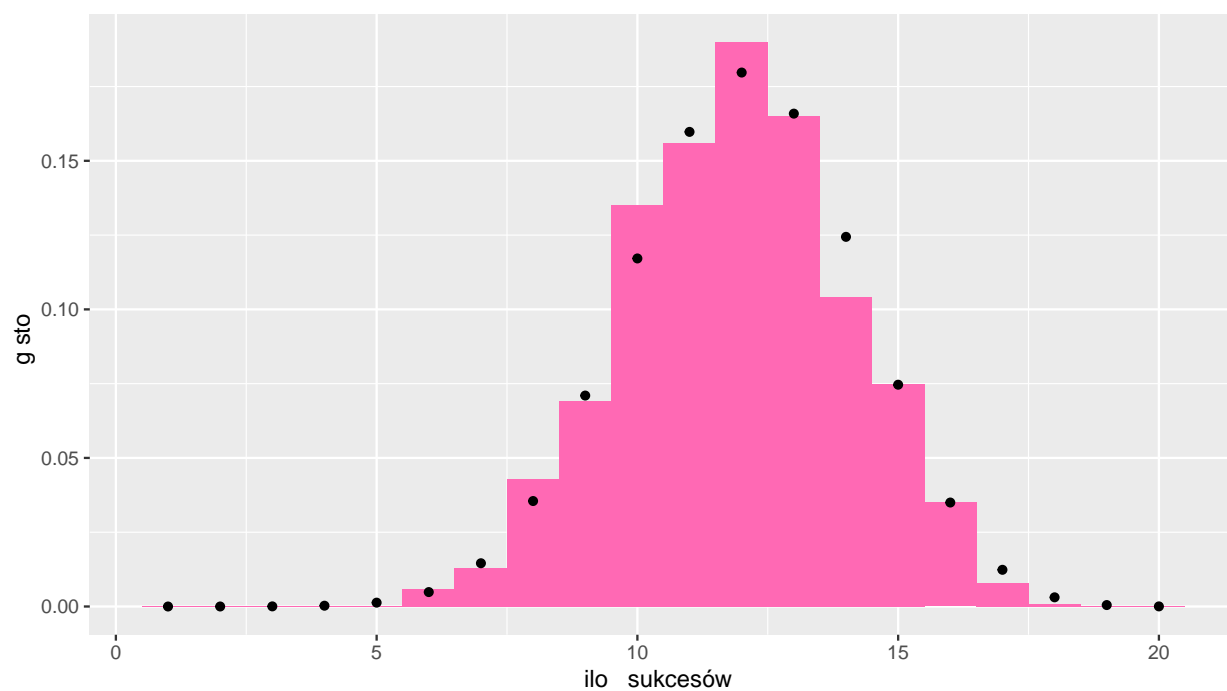
## Porównanie funkcji

Tab. 26: Wartości statystyk

funkcja	średnia	wariancja
rbinom	12.077	4.695767
generuj_dwumianowy	11.878	4.587704



Wykres 9: Histogram dla próby otrzymanej przy pomocy funkcji wbudowanej



Wykres 10: Histogram dla próby otrzymanej przy pomocy funkcji generuj\_dwumianowy.

### Generowanie rozkładu dwumianowego (sposób 2)

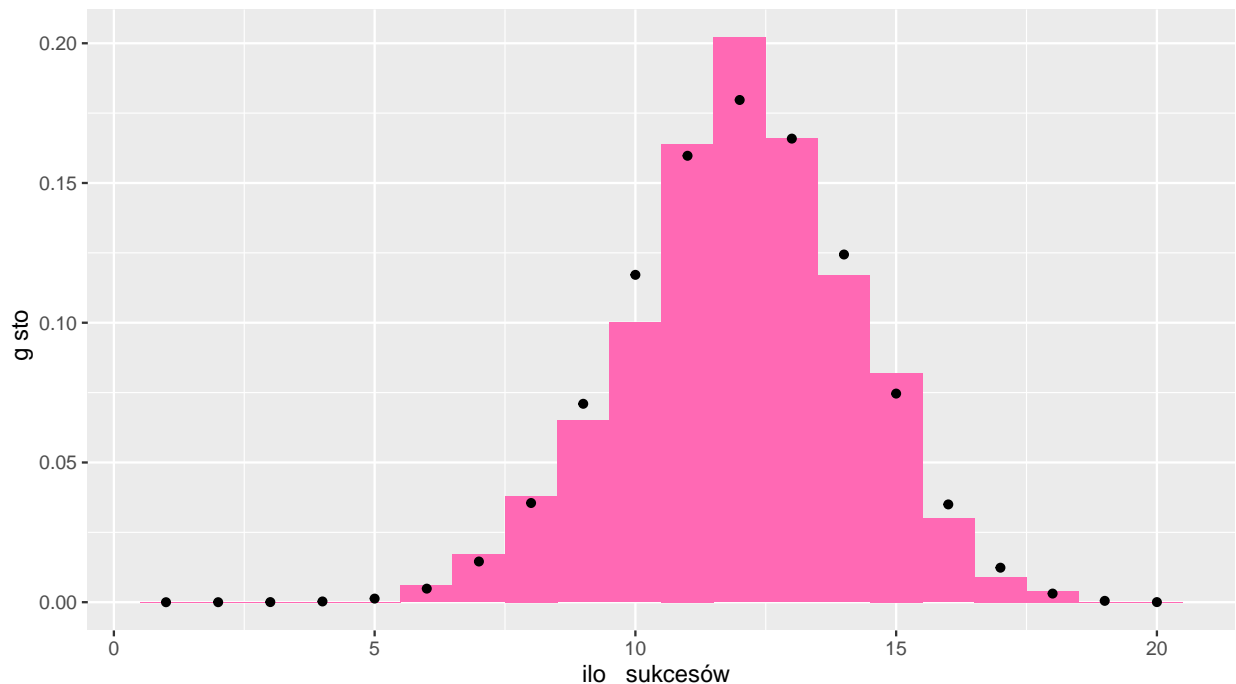
W algorytmie poniżej generujemy rozkład dwumianowy za pomocą rozkładu jednostajnego.

Algorytm ten opiera się na definicji rozkładu dwumianowego, który jest dyskretnym rozkładem prawdopodobieństwa, opisującym liczbę sukcesów w ustalonej liczbie niezależnych prób ze stałym prawdopodobieństwem sukcesu.

W algorytmie symulujemy dużą liczbę prób ze stałą liczbą prób ( $n$ ) i stałym prawdopodobieństwem sukcesu ( $p$ ). Dla każdej próby generujemy losową liczbę z przedziału od 0 do 1 (standardowy rozkład jednostajny) i porównujemy ją z prawdopodobieństwem sukcesu,  $p$ . Jeśli liczba losowa jest mniejsza lub równa  $p$ , liczymy to jako sukces; w przeciwnym razie traktujemy to jako porażkę. Powtarzamy ten proces  $n$  razy, aby zasymulować pojedynczy rozkład dwumianowy, a cały proces powtarzamy wiele razy, aby wygenerować próbkę rozkładów dwumianowych. Uzyskujemy próbkę rozkładów dwumianowych, która jest zbliżona do rzeczywistego rozkładu dwumianowego.

```
n <- 20
p <- 0.6
liczba_probek <- 1:1000
sukcesy <- numeric(length(liczba_probek))

for (i in liczba_probek) {
  próby <- runif(n)
  sukcesy[i] <- sum(próby <= p)
}
```



Wykres 11: Histogram dla próby otrzymanej przy pomocy rozkładu jednostajnego.

## Zadanie 9

### Opis symulacji

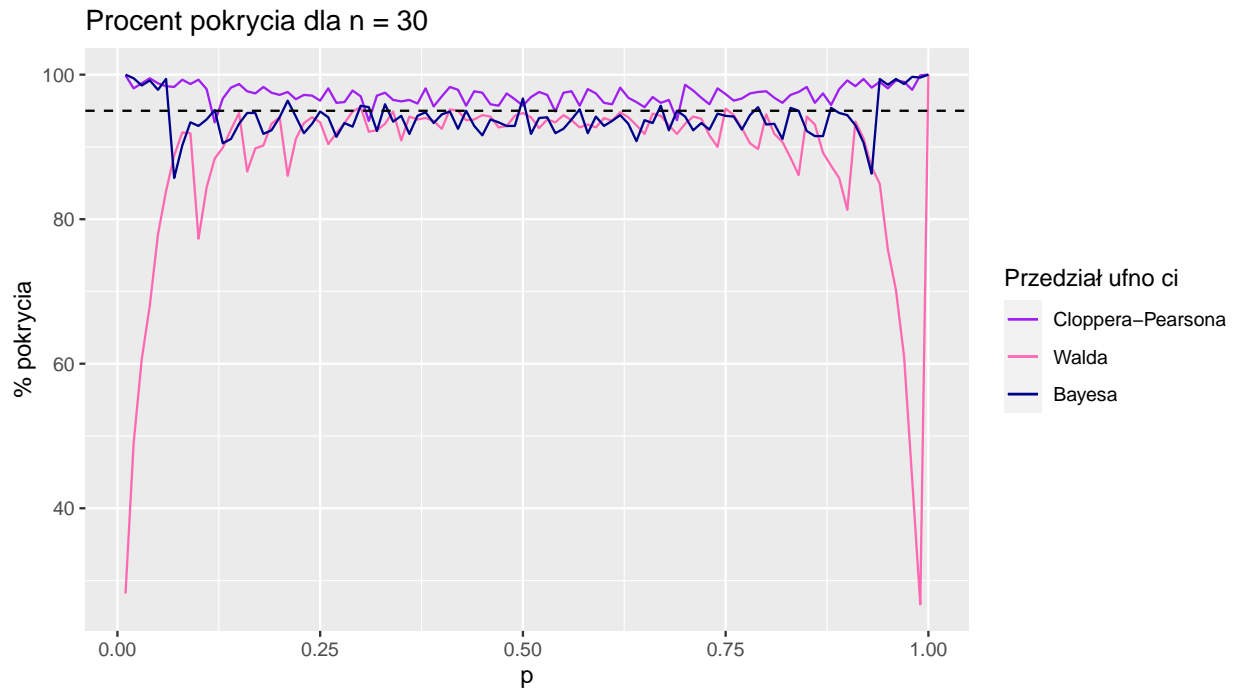
Przeprowadzimy symulację bazującą na metodzie Monte Carlo. Jej celem jest porównanie prawdopodobieństwa pokrycia oraz długości następujących przedziałów ufności: Cloppera-Pearsona, Walda

i Bayesa w zależności od rozmiaru próby  $n \in \{30, 100, 1000\}$  i wartości prawdopodobieństwa  $p \in [0, 1]$ . Symulacja opiera się na poniższym algorytmie.

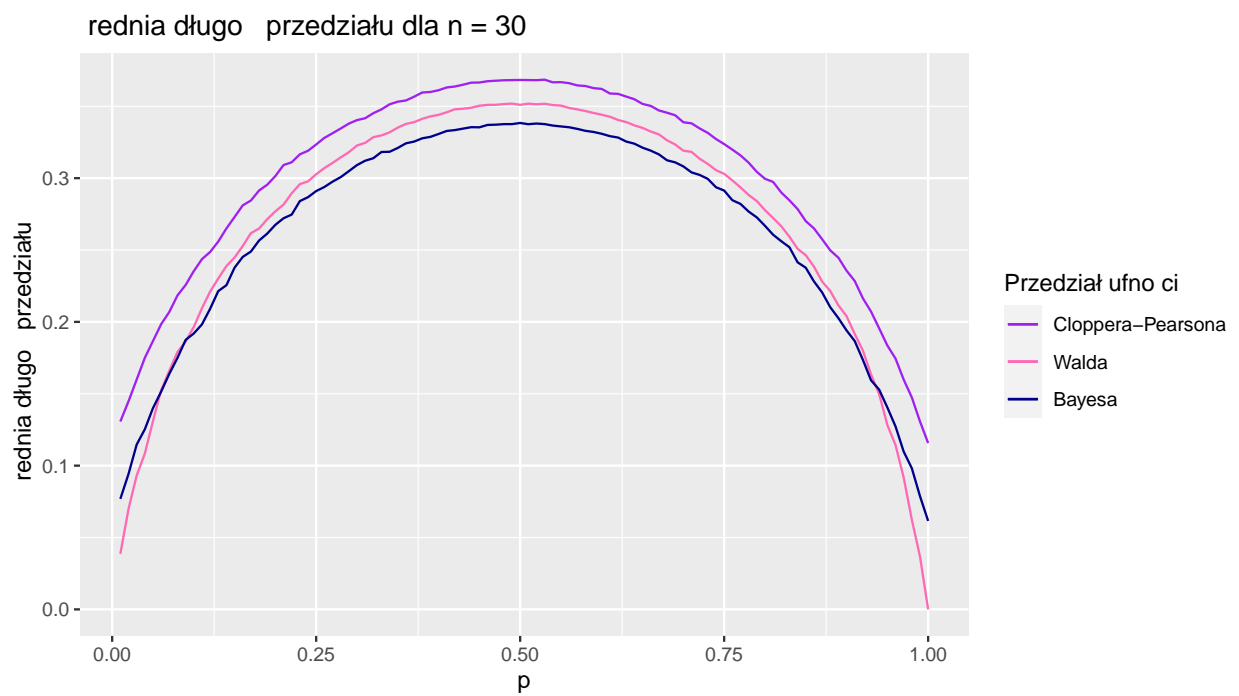
### Algorytm

1. Ustal  $n$  i  $p$ .
2. Generuj realizację zmiennej losowej z rozkładu  $\mathcal{B}(n, p)$ .
3. Wyznacz przedział ufności dla parametru  $p$  wybraną metodą.
4. Sprawdź czy  $p \in$  przedziału ufności. Wyznacz długość przedziału.
5. Powtórz 2-4  $N$  razy.
6. Wyznacz procent pokrycia i średnią długość przedziału.

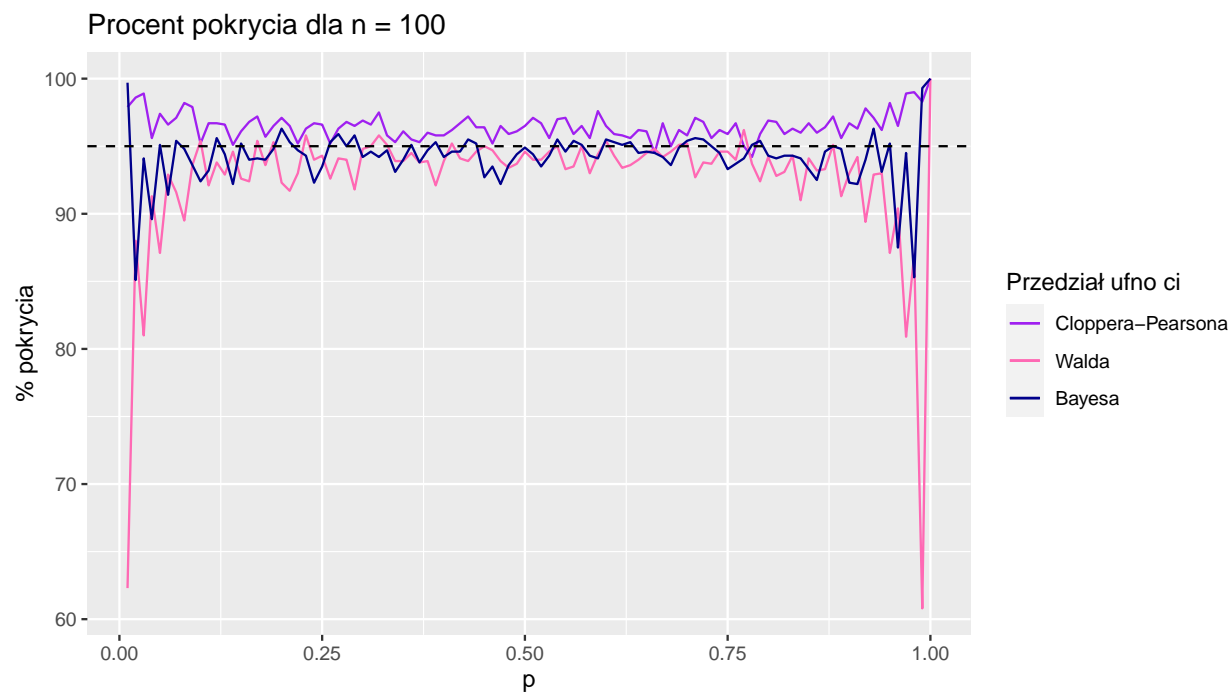
```
simulation <- function(n, method, name){
  N <- 1000
  p <- seq(0, 1, 0.01)
  for(j in 1:length(p)){
    len <- rep(NA, N)
    counter <- 0
    for(i in 1:N){
      x <- rbinom(1, n, p[j])
      interval <- binom.confint(x, n, conf.level = 0.95, methods = method)
      TL <- interval$lower
      TU <- interval$upper
      if(between(p[j], TL, TU)){
        counter <- counter + 1
      }
      len[i] <- TU - TL
    }
    p.cover <- counter/N*100
    p.len <- mean(len)
    df <- data.frame('p' = p[j], 'pokrycie' = p.cover, 'dlugosc' = p.len)
    write.table(df, name,
      append = TRUE,
      sep = ",",
      col.names = FALSE,
      row.names = FALSE,
      quote = FALSE)
  }
}
```



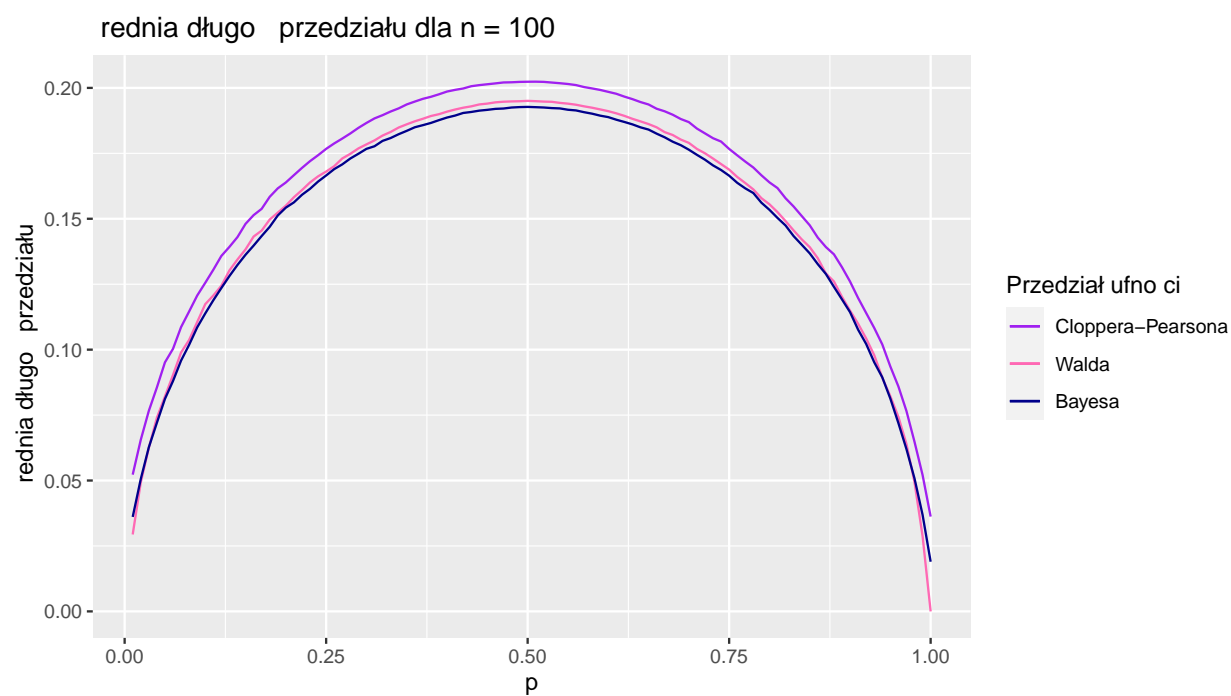
Wykres 12: Procent pokrycia dla  $n = 30$



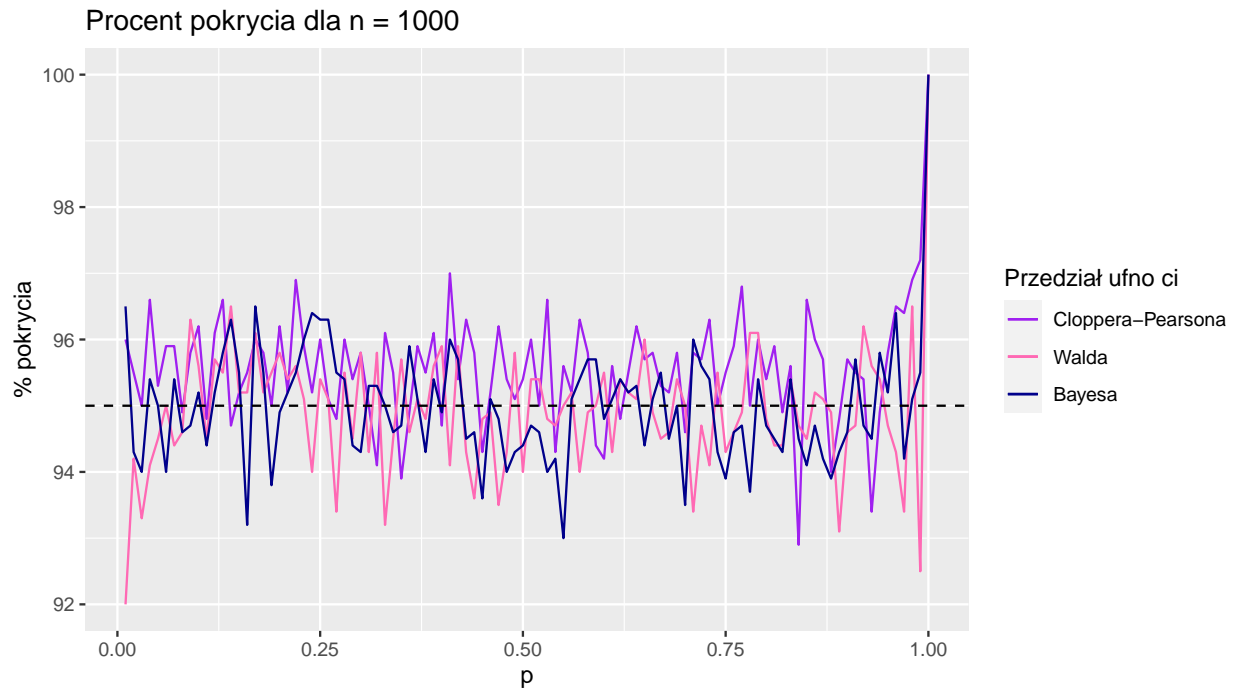
Wykres 13: Średnia długość przedziału dla  $n = 30$



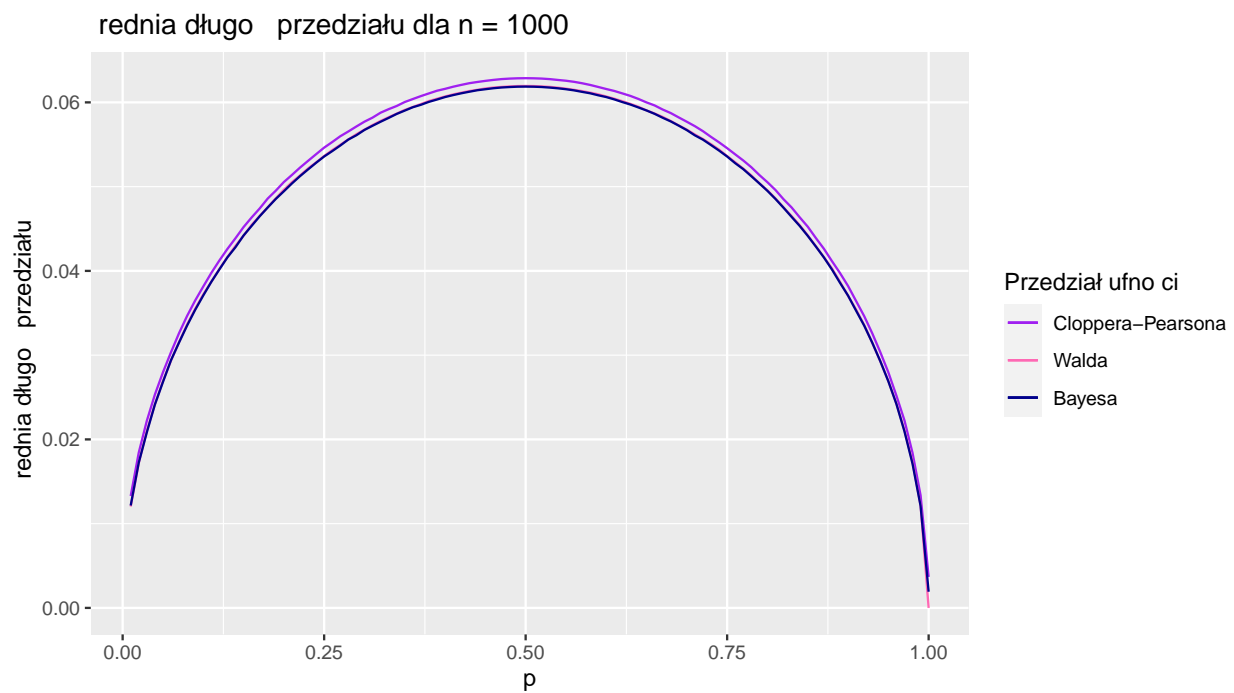
Wykres 14: Procent pokrycia dla  $n = 100$



Wykres 15: Średnia długość przedziału dla  $n = 100$



Wykres 16: Procent pokrycia dla  $n = 1000$



Wykres 17: Średnia długość przedziału dla  $n = 1000$

## Wnioski

W przypadku wykresów 12 i 14 dla  $n = 30$  i  $n = 100$  możemy jednoznacznie stwierdzić, że największy procent pokrycia uzyskujemy dla przedziałów ufności Cloppera-Pearsona, co nie dziwi biorąc pod uwagę średnią

długość przedziałów (wykresy 13 i 15), ponieważ właśnie te przedziały są znacząco szersze od pozostałych. W tym przypadku najgorzej wypadają przedziały ufności Walda, ponieważ przy zbliżonej długości przedziału do przedziałów Bayesa dla  $n = 100$  dają gorszy procent pokrycia, szczególnie na krańcach przedziałów.

Dla  $n = 1000$  procent pokrycia oscyluje w granicy 95% dla wszystkich przedziałów (wykres 16). Nieznacznie lepiej na krańcach wypadają przedziały Cloppera-Pearsona oraz Bayesa. Również średnia długość przedziałów jest zbliżona, minimalnie szersze są ponownie przedziały Cloppera-Pearsona.

## Część IV

### Zadanie 11

#### Weryfikacja hipotez

Wszystkie poniższe hipotezy są weryfikowane na poziomie istotności 0.05.

- a) Testujemy hipotezę zerową  $H_0$  : prawdopodobieństwo, że w korporacji pracuje kobieta wynosi 0.5, przeciwko hipotezie alternatywnej  $H_1$  : prawdopodobieństwo, że w korporacji pracuje kobieta jest różne od 0.5.

- Wyniki uzyskane przy pomocy funkcji `binom.test`

parametr	wartość
statistic.number of successes	71
parameter.number of trials	200
p.value	4.97297347921349e-05
conf.int1	0.288783777713584
conf.int2	0.42558616692658
estimate.probability of success	0.355
null.value.probability of success	0.5
alternative	two.sided
method	Exact binomial test

- Wyniki uzyskane przy pomocy funkcji `prop.test` z uwzględnieniem poprawki na ciągłość

parametr	wartość
statistic.X-squared	16.245
parameter.df	1
p.value	5.56562799613989e-05
estimate.prop 1	0.355
estimate.prop 2	0.5
conf.int1	0.289636295915694
conf.int2	0.42603274730263
alternative	two.sided
method	1-sample proportions test with continuity correction

- Wyniki uzyskane przy pomocy funkcji `prop.test` bez uwzględnienia poprawki na ciągłość



parametr	wartość
statistic.X-squared	16.82
parameter.df	1
p.value	4.10978780994588e-05
estimate.prop 1	0.355
estimate.prop 2	0.5
conf.int1	0.291986275404312
conf.int2	0.423478869245096
alternative	two.sided
method	1-sample proportions test without continuity correction

### Wnioski

Korzystając z testu dokładnego (`binom.test`) na poziomie istotności  $\alpha = 0.05$  weryfikowaną hipotezę, że prawdopodobieństwo, iż w korporacji pracuje kobieta wynosi 0.5 należy odrzucić (p-wartość w tym teście wynosi  $4.9729735 \times 10^{-5}$ ). Podobnie w przypadku testu asymptotycznego (`prop.test`) na poziomie istotności  $\alpha = 0.05$  z zastosowaną poprawką na ciągłość (p-wartość wynosi  $5.565628 \times 10^{-5}$ ) oraz bez uwzględniania poprawki (p-wartość  $4.1097878 \times 10^{-5}$ ) również odrzucamy weryfikowaną hipotezę na rzecz hipotezy alternatywnej, że prawdopodobieństwo, iż w korporacji pracuje kobieta jest różne od 0.5.

- b) Testujemy hipotezę zerową  $H_0$  : prawdopodobieństwo, że pracownik jest zadowolony ze swojego wynagrodzenia jest większe bądź równe 0.8, przeciwko hipotezie alternatywnej  $H_1$  : prawdopodobieństwo, że pracownik jest zadowolony ze swojego wynagrodzenia jest mniejsze od 0.8.

- Wyniki uzyskane przy pomocy funkcji `binom.test`

parametr	wartość
statistic.number of successes	106
parameter.number of trials	200
p.value	9.25478699944125e-18
conf.int1	0
conf.int2	0.589919355385537
estimate.probability of success	0.53
null.value.probability of success	0.8
alternative	less
method	Exact binomial test

- Wyniki uzyskane przy pomocy funkcji `prop.test` z poprawką na ciągłość

parametr	wartość
statistic.X-squared	89.4453125
parameter.df	1
p.value	1.57616542390591e-21
estimate.prop 1	0.53
estimate.prop 2	0.8
conf.int1	0
conf.int2	0.589710647995846
alternative	less
method	1-sample proportions test with continuity correction

- Wyniki uzyskane przy pomocy funkcji `prop.test` bez poprawki na ciągłość

parametr	wartość
statistic.X-squared	91.125
parameter.df	1
p.value	6.7438394468056e-22
estimate.prop 1	0.53
estimate.prop 2	0.8
conf.int1	0
conf.int2	0.587261860986402
alternative	less
method	1-sample proportions test without continuity correction

### Wnioski

Korzystając z testu dokładnego (**binom.test**) na poziomie istotności  $\alpha = 0.05$  weryfikowaną hipotezę, że prawdopodobieństwo, iż pracownik jest zadowolony ze swojego wynagrodzenia jest większe bądź równe 0.8 należy odrzucić (p-wartość w tym teście wynosi  $9.254787 \times 10^{-18}$ ). Podobnie w przypadku testu asymptotycznego (**prop.test**) na poziomie istotności  $\alpha = 0.05$  z zastosowaną poprawką na ciągłość (p-wartość wynosi  $1.5761654 \times 10^{-21}$ ) oraz bez uwzględniania poprawki (p-wartość  $6.7438394 \times 10^{-22}$ ) również odrzucamy weryfikowaną hipotezę na rzecz hipotezy alternatywnej, że prawdopodobieństwo, iż pracownik jest zadowolony ze swojego wynagrodzenia jest mniejsze od 0.8.

- c) Testujemy hipotezę, że prawdopodobieństwo, iż kobieta pracuje na stanowisku kierowniczym jest równe prawdopodobieństwu, że mężczyzna pracuje na stanowisku kierowniczym.

Wyniki uzyskane przy pomocy funkcji **prop.test**

parametr	wartość
statistic.X-squared	0.220136236322974
parameter.df	1
p.value	0.638936132739115
estimate.prop 1	0.112676056338028
estimate.prop 2	0.147286821705426
conf.int1	-0.141181729728431
conf.int2	0.0719601989936348
alternative	two.sided
method	2-sample test for equality of proportions with continuity correction

### Wnioski

Korzystając z testu asymptotycznego (**prop.test**) na poziomie istotności  $\alpha = 0.05$  nie ma podstaw do odrzucenia hipotezy, że prawdopodobieństwo, iż kobieta pracuje na stanowisku kierowniczym jest równe prawdopodobieństwu, że mężczyzna pracuje na stanowisku kierowniczym. Wartość poziomu krytycznego wynosi 0.6389361.

- d) Testujemy hipotezę, że prawdopodobieństwo, iż kobieta jest zadowolona ze swojego wynagrodzenia jest równe prawdopodobieństwu, że mężczyzna jest zadowolony ze swojego wynagrodzenia.

parametr	wartość
statistic.X-squared	0.111934791869356
parameter.df	1
p.value	0.737952063846349
estimate.prop 1	0.507042253521127
estimate.prop 2	0.542635658914729
conf.int1	-0.191128833124416
conf.int2	0.119942022337212
alternative	two.sided
method	2-sample test for equality of proportions with continuity correction

### Wnioski

Korzystając z testu asymptotycznego (**prop.test**) na poziomie istotności  $\alpha = 0.05$  nie ma podstaw do odrzucenia hipotezy, że prawdopodobieństwo, iż kobieta jest zadowolona ze swojego wynagrodzenia jest równe prawdopodobieństwu, że mężczyzna jest zadowolony ze swojego wynagrodzenia. Wartość poziomu krytycznego wynosi 0.7379521.

- e) Testujemy hipotezę, że prawdopodobieństwo, iż kobieta pracuje w dziale obsługi kadrowo-płacowej jest większe lub równe prawdopodobieństwu, że mężczyzna pracuje w dziale obsługi kadrowo-płacowej.

parametr	wartość
statistic.X-squared	33.9986112329878
parameter.df	1
p.value	0.999999997242429
estimate.prop 1	0.323943661971831
estimate.prop 2	0.0232558139534884
conf.int1	-1
conf.int2	0.405530681459733
alternative	less
method	2-sample test for equality of proportions with continuity correction

### Wnioski

Na podstawie przeprowadzonego testu asymptotycznego (**prop.test**) nie mamy podstaw do odrzucenia hipotezy, że prawdopodobieństwo, iż kobieta pracuje w dziale obsługi kadrowo-płacowej jest większe lub równe prawdopodobieństwu, że mężczyzna pracuje w dziale obsługi kadrowo-płacowej. Wartość poziomu krytycznego w tym teście wynosi w przybliżeniu 1.