

Analiza danych ankietowych

Sprawozdanie 1

Zuzanna Nasiłowska Maria Nowacka

Spis treści

1 Część 1.

1.1 Zadanie 1

W pewnej dużej firmie technologicznej przeprowadzono ankietę mającą na celu ocenę skuteczności programów szkoleniowych dla pracowników. Wzięło w niej udział 200 losowo wybranych osób (losowanie proste ze zwracaniem).

1.1.1 Zadanie 1.1

Wczytamy dane.

```
[1] 200    8
```

Dane zawierają 200 wierszy oraz 8 kolumn.

Sprawdzamy typy zmiennych.

DZIAŁ	STAŻ	CZY_KIER	PYT_1	PYT_2	PYT_3
"character"	"integer"	"character"	"integer"	"integer"	"integer"
PŁEĆ	WIEK				
"character"	"integer"				

Wszystkie zmienne o typie *character* przekształcamy na typ *factor*.

Liczba wartości brakujących wynosi: 0

Sprawdzamy, czy typy zmiennych zostały prawidłowo rozpoznane.

1. Zmienne ilościowe (typ numeric)

STAŻ	PYT_1	PYT_2	PYT_3	WIEK
2	4	5	6	8

2. Zmienne jakościowe (typ factor)

DZIAŁ	CZY_KIER	PŁEĆ
1	3	7

1.1.2 Zadanie 1.2

Utwórz zmienną WIEK_KAT przeprowadzając kategoryzację zmiennej WIEK korzystając z następujących przedziałów: do 35 lat, między 36 a 45 lat, między 46 a 55 lat, powyżej 55 lat.

1.1.3 Zadanie 1.3

Sporządź tablice liczości dla zmiennych: DZIAŁ, STAŻ, CZY_KIER, PŁEĆ, WIEK_KAT. Sformułuj wnioski.

DZIAŁ

HR IT MK PD

31 26 45 98

STAŻ

1 2 3

41 140 19

CZY_KIER

Nie Tak

173 27

PŁEĆ

K M

71 129

WIEK_KAT

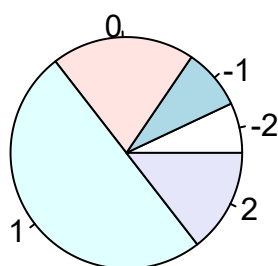
<35 36-40 46-55 >55

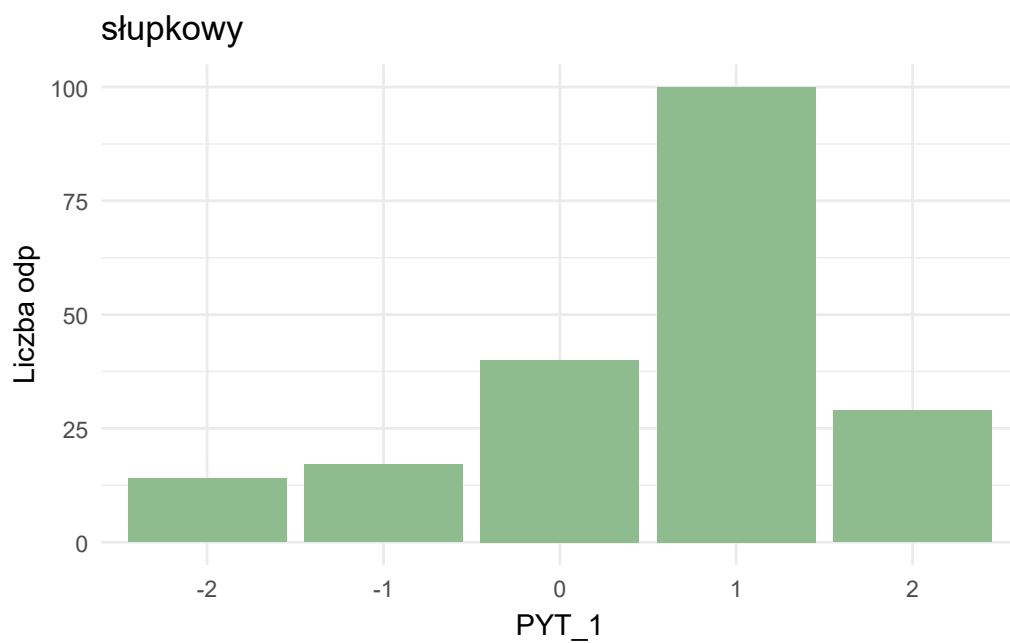
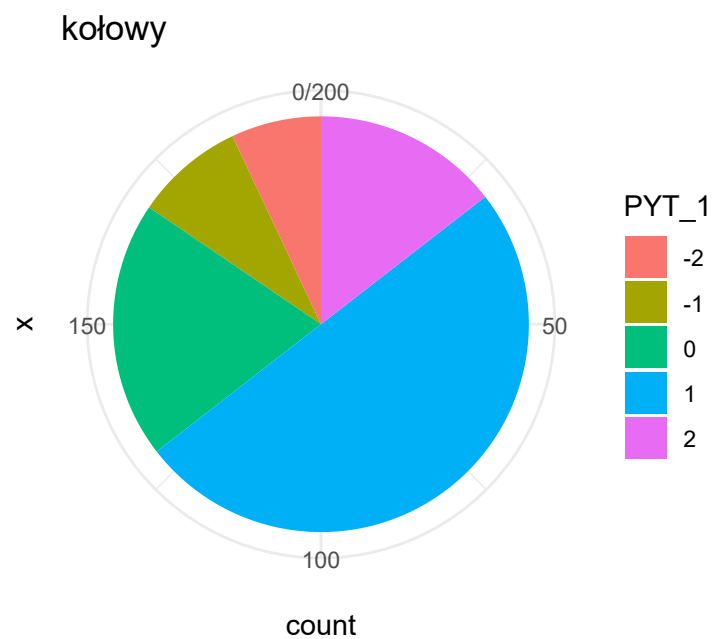
26 104 45 25

Na podstawie tabel liczości możemy zauważyć, że: - W firmie prawie połowa pracowników jest zatrudniona w dziale “PD” (Dział Produktowy). Drugi największy dział to “MK” (Marketing), następnie “HR” (Dział zasobów ludzkich). Najmniej pracowników jest zatrudnionych w dziale “IT”. - Najwięcej osób pracuje w firmie między jednym a trzema latami. Mało osób ma staż ponad 3 lata. - W firmie 27 osób ma stanowisko kierownicze. - Większość pracowników to mężczyźni. - Ponad połowa pracowników jest w wieku 36-40 lat.

1.1.4 Zadanie 1.4

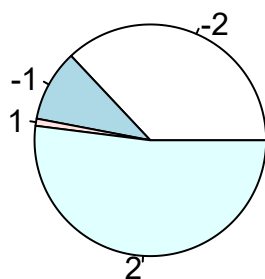
Sporządź wykresy kołowe oraz wykresy słupkowe dla zmiennych: PYT_1 oraz PYT_2. Sformułuj wnioski.



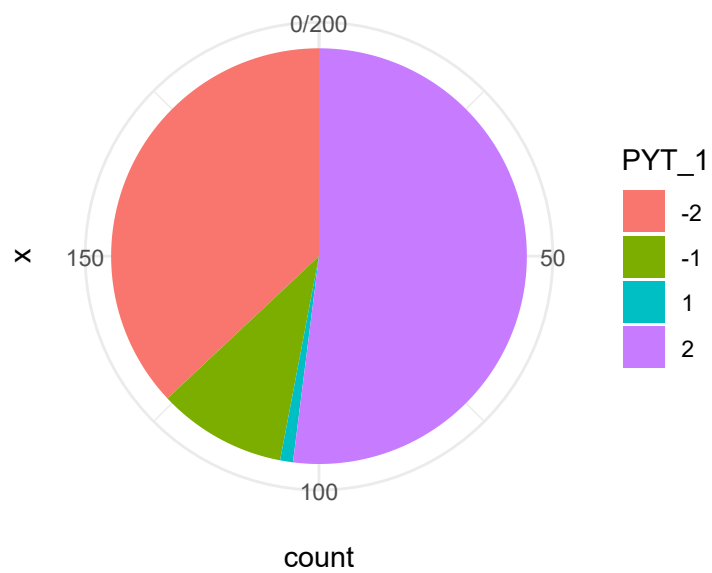


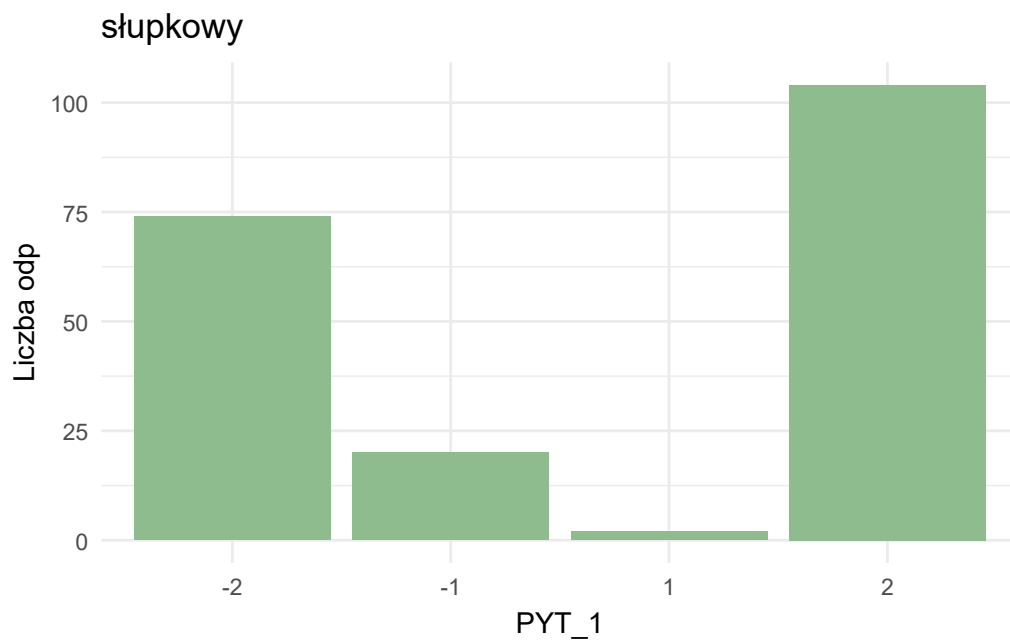
Na pytanie “Jak bardzo zgadzasz się ze stwierdzeniem, że firma zapewnia odpowiednie wsparcie i materiały umożliwiające skuteczne wykorzystanie w

praktyce wiedzy zdobytej w trakcie szkoleń?”



kołowy





1.1.5 Zadanie 1.5

1.2 Zadanie 6

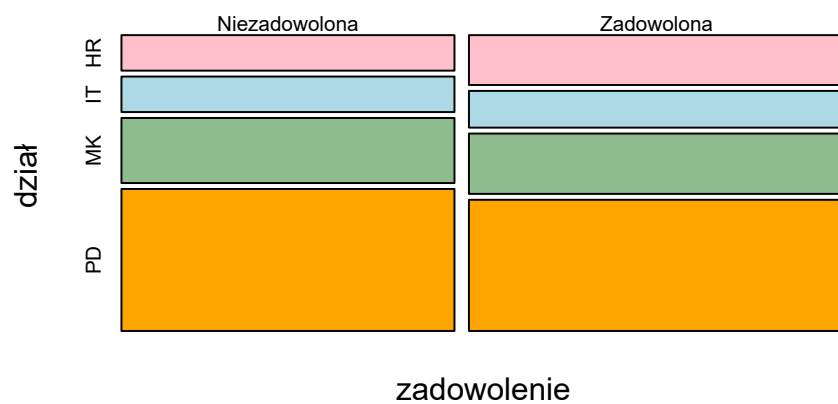
	PYT_3			
PYT_2	-2	-1	1	2
-2	49	16	5	4
-1	3	6	10	1
1	0	0	2	0
2	0	8	15	81

1.3 Zadanie 7

```
#ankieta$CZY_ZADOW <- ifelse(ankieta$PYT_2 %in% c(-2, -1), "Nie zgadzam się",
#                               ifelse(ankieta$PYT_2 %in% c(1, 2), "Zgadzam się", NA),
#
```

1.4 Zadanie 8

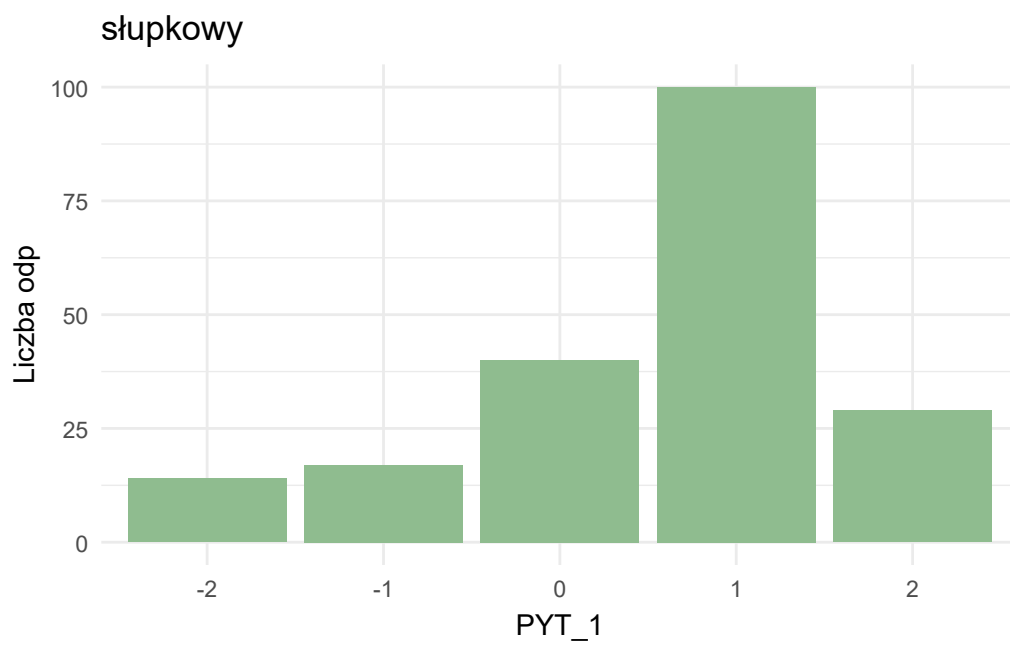
```
table(ankieta$CZY_ZADOW, ankieta$DZIAŁ)
```



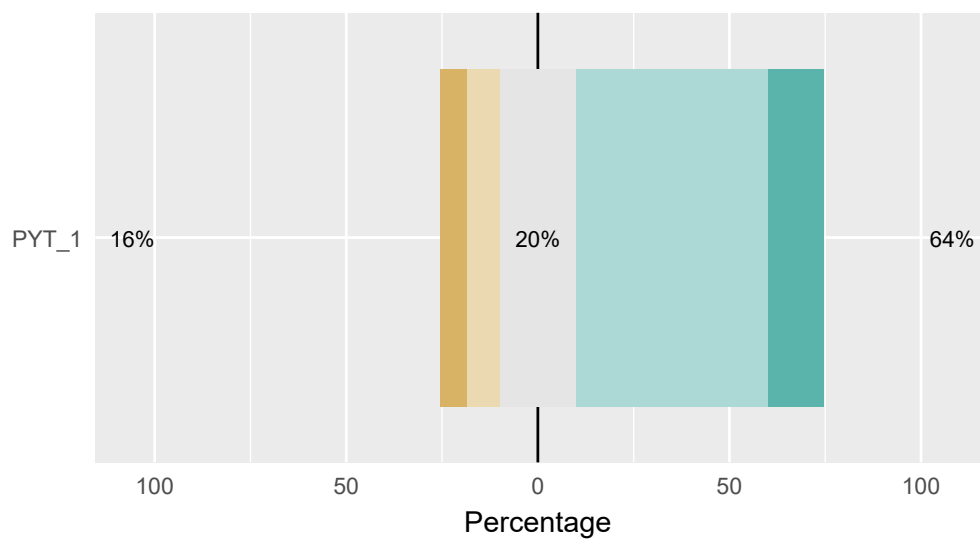
1.5 Część 2.

1.6 zadanie 2

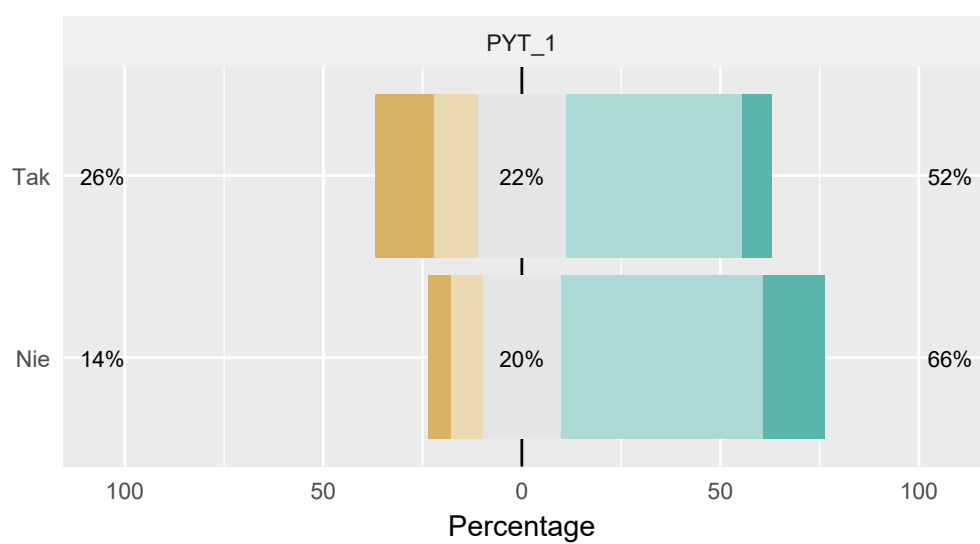
```
[1] "factor"
```

Group	Item	-2	-1	0	1	2
1	Nie PYT_1	5.780347	8.092486	19.65318	50.86705	15.606936
2	Tak PYT_1	14.814815	11.111111	22.22222	44.44444	7.407407

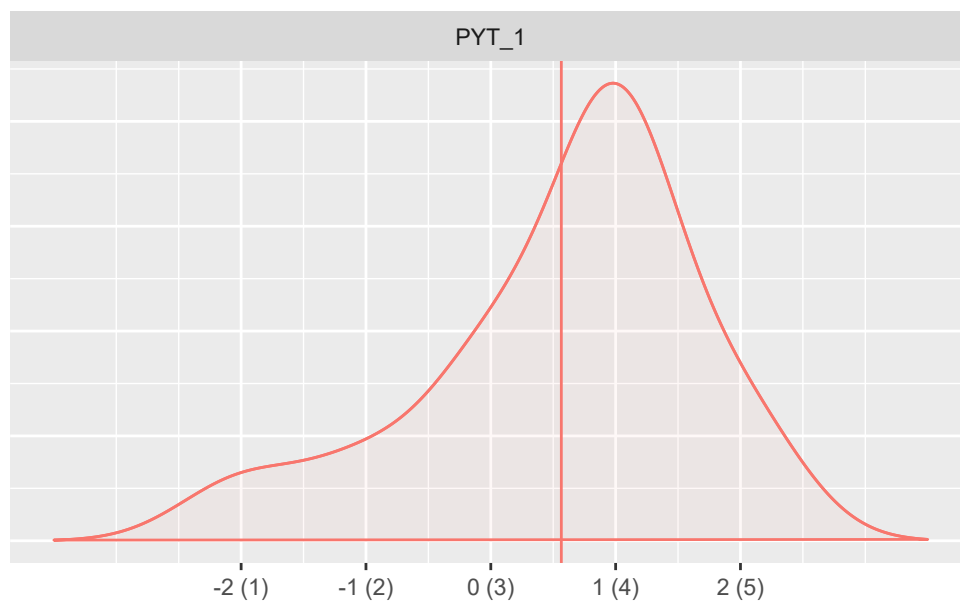


Response -2 -1 0 1 2



Response -2 -1 0 1 2

-2 -1 0 1 2
14 17 40 100 29



1.7 Zadanie 3

```
library(stats)
#ankieta[sample(20),]
ankieta[sample(1:nrow(ankieta), size = 0.1*nrow(ankieta), replace = FALSE),]
```

	DZIAŁ	STAŻ	CZY_KIER	PYT_1	PYT_2	PYT_3	PŁEĆ	WIEK	WIEK_KAT	CZY_ZADOW
189	HR	2	Nie	1	2	2	M	49	46-55	Zadowolona
115	PD	1	Nie	-2	-2	-2	M	39	36-40	Niezadowolona
8	IT	2	Tak	1	-1	-2	K	58	>55	Niezadowolona
141	MK	2	Nie	1	2	2	K	45	36-40	Zadowolona
110	PD	2	Nie	0	-2	2	K	45	36-40	Niezadowolona
192	HR	2	Nie	0	-2	-1	M	64	>55	Niezadowolona
72	PD	2	Nie	2	2	2	M	33	<35	Zadowolona
172	HR	2	Nie	2	2	2	M	70	>55	Zadowolona
160	MK	2	Nie	1	2	2	M	30	<35	Zadowolona
21	IT	2	Nie	1	-1	1	K	60	>55	Niezadowolona

32	PD	1	Nie	0	-2	-2	M	33	<35	Niezadowolona
124	PD	2	Nie	0	-2	-2	M	40	36-40	Niezadowolona
116	PD	1	Nie	1	-1	-1	M	37	36-40	Niezadowolona
69	PD	1	Nie	-1	-2	-1	M	25	<35	Niezadowolona
140	MK	2	Nie	2	2	-1	K	38	36-40	Zadowolona
148	MK	2	Nie	1	2	2	K	48	46-55	Zadowolona
125	MK	2	Nie	1	2	2	K	40	36-40	Zadowolona
107	PD	2	Nie	0	-2	-2	K	43	36-40	Niezadowolona
149	MK	1	Nie	0	-2	-1	M	42	36-40	Niezadowolona
187	HR	2	Tak	2	2	2	M	48	46-55	Zadowolona

```
ankieta[sample(1:nrow(ankieta), size = 0.1*nrow(ankieta), replace = TRUE),]
```

	DZIAŁ	STAŻ	CZY_KIER	PYT_1	PYT_2	PYT_3	PŁEĆ	WIEK	WIEK_KAT	CZY_ZADOW
200	HR	2	Nie	1	2	2	K	42	36-40	Zadowolona
104	PD	3	Tak	1	2	2	K	36	36-40	Zadowolona
117	PD	2	Nie	1	2	2	M	40	36-40	Zadowolona
59	PD	2	Nie	1	-2	-2	M	49	46-55	Niezadowolona
199	HR	2	Nie	1	2	1	K	48	46-55	Zadowolona
23	IT	2	Nie	-1	-2	-2	K	60	>55	Niezadowolona
99	PD	2	Nie	1	2	2	M	40	36-40	Zadowolona
81	PD	2	Nie	1	2	-1	M	44	36-40	Zadowolona
133	MK	2	Nie	1	2	2	K	39	36-40	Zadowolona
164	MK	2	Nie	0	-2	-2	M	54	46-55	Niezadowolona
69	PD	1	Nie	-1	-2	-1	M	25	<35	Niezadowolona
138	MK	2	Nie	2	2	2	K	39	36-40	Zadowolona
137	MK	2	Nie	2	2	2	K	41	36-40	Zadowolona
146	MK	3	Nie	1	2	2	K	52	46-55	Zadowolona
32	PD	1	Nie	0	-2	-2	M	33	<35	Niezadowolona
197	HR	2	Nie	1	-1	-1	K	35	<35	Niezadowolona
176	HR	2	Nie	1	-1	-1	M	40	36-40	Niezadowolona
36	PD	1	Nie	-2	-2	1	M	29	<35	Niezadowolona
57	PD	2	Nie	1	2	2	M	47	46-55	Zadowolona
151	MK	3	Nie	-2	-2	1	M	40	36-40	Niezadowolona

1.8 zadanie 4

```
symulacja <- function(N,n, p) {  
  
  wyniki <- numeric(N)  
  
  for(i in 1:N) {  
    bernoulli <- rbinom(n = n, size = 1, prob = p)  
    wyniki[i] <- sum(bernoulli)  
  }  
  
  return(wyniki)  
}  
n <- 200  
p <- 0.2  
N <- 10000  
  
samples_1 <- symulacja(N, n, p)  
wartość_oczekiwana_emp <- mean(samples_1)  
war_emp <- var(samples_1)  
  
wartość_oczekiwana_teo <- n * p  
wariancja_teo <- n * p *(1 - p)  
  
cat("Teoretyczna wartość oczekiwana: ", wartość_oczekiwana_teo, "\n")
```

Teoretyczna wartość oczekiwana: 40

```
cat("Teoretyczna wariancja: ", wariancja_teo, "\n")
```

Teoretyczna wariancja: 32

```
cat("empiryczna wartość oczekiwana: ", wartość_oczekiwana_emp, "\n")
```

empiryczna wartość oczekiwana: 40.0923

```
cat("empiryczna wariancja: ", war_emp, "\n")
```

```
empiryczna wariancja: 31.75096
```

1.9 zadanie 5

```
ps <- c(0.1, 0.23, 0.47, 0.17, 0.03)
k <- length(ps)
sum(ps)
```

```
[1] 1
```

```
N <- 10000
csum = cumsum(ps)
X <- rep(0, k)
for (i in 1:N){
  Z <- runif(1)
  for (j in 1:k){
    if (Z<csum[j]){
      X[j] <- X[j] + 1
      break
    }
  }
}
X/N
```

```
[1] 0.0990 0.2287 0.4712 0.1699 0.0312
```

```
ps2 <- rep(0.2, 5)
k <- length(ps2)
sum(ps)
```

```
[1] 1
```

```

N <- 10000
csum2 = cumsum(ps2)
Y <- rep(0, k)
for (i in 1:N){
  Z <- runif(1)
  for (j in 1:k){
    if (Z<csum2[j]){
      Y[j] <- Y[j] + 1
      break
    }
  }
}
Y/N

```

```
[1] 0.1961 0.2034 0.1981 0.1998 0.2026
```