

Lista 2

Mikołaj Słupiński

24 marca 2018

Zadanie 1

```
tab = read.table("http://www.math.uni.wroc.pl/~mbogdan/Podstawy/Dane/individuals.dat")
set.seed(1337)
at_least_bachelor <- function(tab) {
  return(round(length(tab[tab[[3]] >= 5,3])/length(tab[[3]]) * 100))
}
income_mean = mean(tab[[5]])
print(paste("Średnia dochodów: ", income_mean))
income_median = median(tab[[5]])
print(paste("Mediana dochodów: ", income_median))
income_sd = sd(tab[[5]])
print(paste("Odchylenie standardowe dochodów: ", income_sd))
income_iqr = IQR(tab[[5]])
print(paste("IQR dochodów:", income_iqr))
at_least_bachelor_percent = at_least_bachelor(tab)
print(paste("Procent osób z co najmniej licencjatem: ", at_least_bachelor_percent))
income_left = income_mean - 2 * income_sd
income_right = income_mean + 2 * income_sd
print(paste("Procent zarobków w zadanym przedziale: ",
  round(length(tab[tab[[5]] >= income_left &
    tab[[5]] <= income_right,5]) / length(tab[[5]]) * 100)))
```

```
## [1] "Średnia dochodów: 37864.6094563409"
## [1] "Mediana dochodów: 29717"
## [1] "Odchylenie standardowe dochodów: 36158.0278116033"
## [1] "IQR dochodów: 29503.5"
## [1] "Procent osób z co najmniej licencjatem: 30"
## [1] "Procent zarobków w zadanym przedziale: 96"
```

Z reguły 3σ 94% zarobków powinno być w przedziale $[\mu - 2\sigma, \mu + 2\sigma]$. Otrzymaliśmy ok. 96 %, więc można uznać, że wynik jest zbliżony do oszacowania.

a)

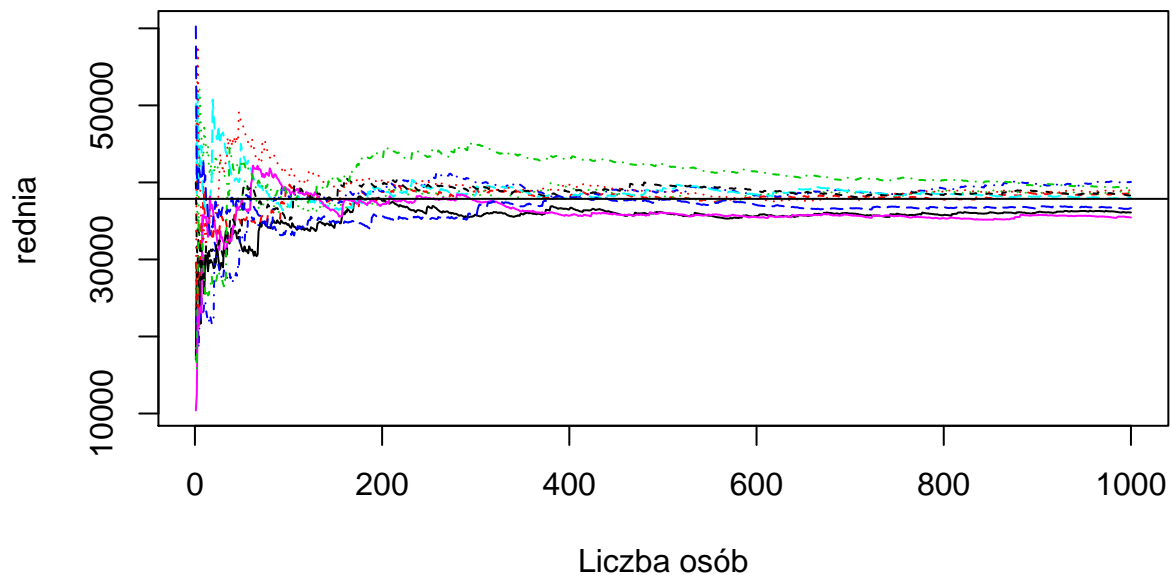
```
par(mfrow=c(1,1))
u = array(0, c(10, 5, 1000))
for(i in 1:10)
{
  indices = sample(dim(tab)[1], 1000)
  s = tab[indices, 5]
  u[i,,] = sapply(1:1000, function(i) c(mean(s[1:i]),
    median(s[1:i]), IQR(s[1:i]),
    sd(s[1:i]), at_least_bachelor(tab[indices,][1:i,])))
}
matplot(t(u[,1,]), type = "l", main = "Trajektorie średniej",
```

```

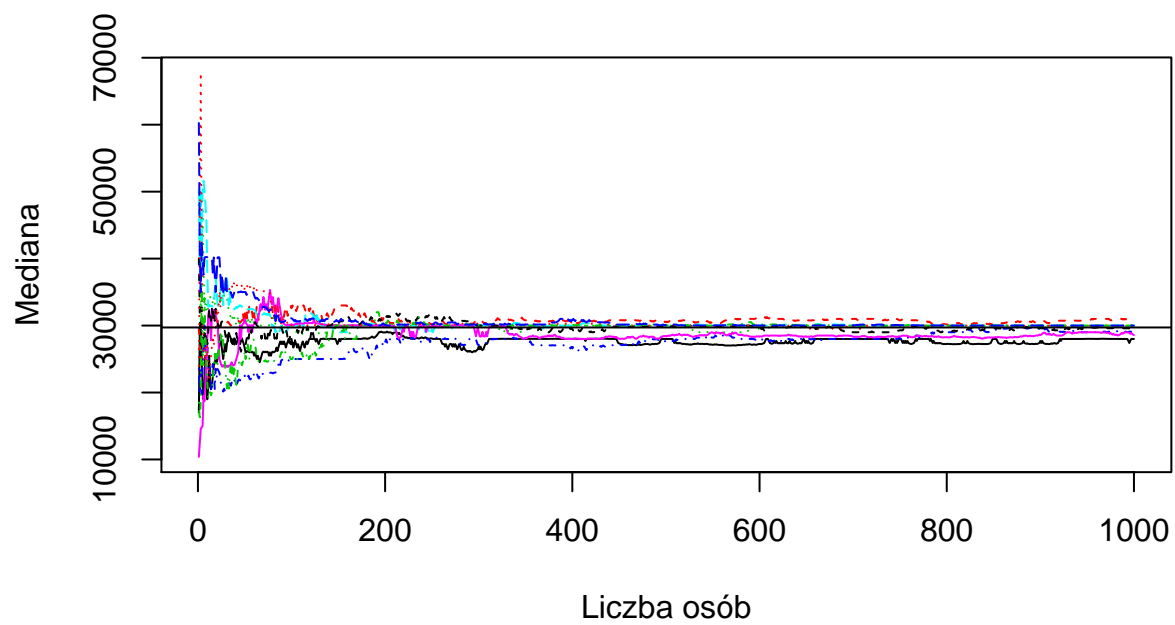
        ylab = "Średnia", xlab = "Liczba osób")
abline(income_mean, 0)
matplot(t(u[,2,]), type = "l", main = "Trajektorie mediany",
        ylab = "Mediana", xlab = "Liczba osób")
abline(income_median, 0)
matplot(t(u[,3,]), type = "l", main = "Trajektorie IQR",
        ylab = "IQR", xlab = "Liczba osób")
abline(income_iqr, 0)
matplot(t(u[,4,]), type = "l", main = "Trajektorie odchylenia standardowego",
        ylab = "Odchylenie standardowe", xlab = "Liczba osób")
abline(income_sd, 0)
matplot(t(u[,5,]), type = "l", main = "Trajektorie procenta osób z conajmniej licencjatem",
        ylab = "Procent osób z conajmniej licencjatem", xlab = "Liczba osób")
abline(at_least_bachelor_percent, 0)

```

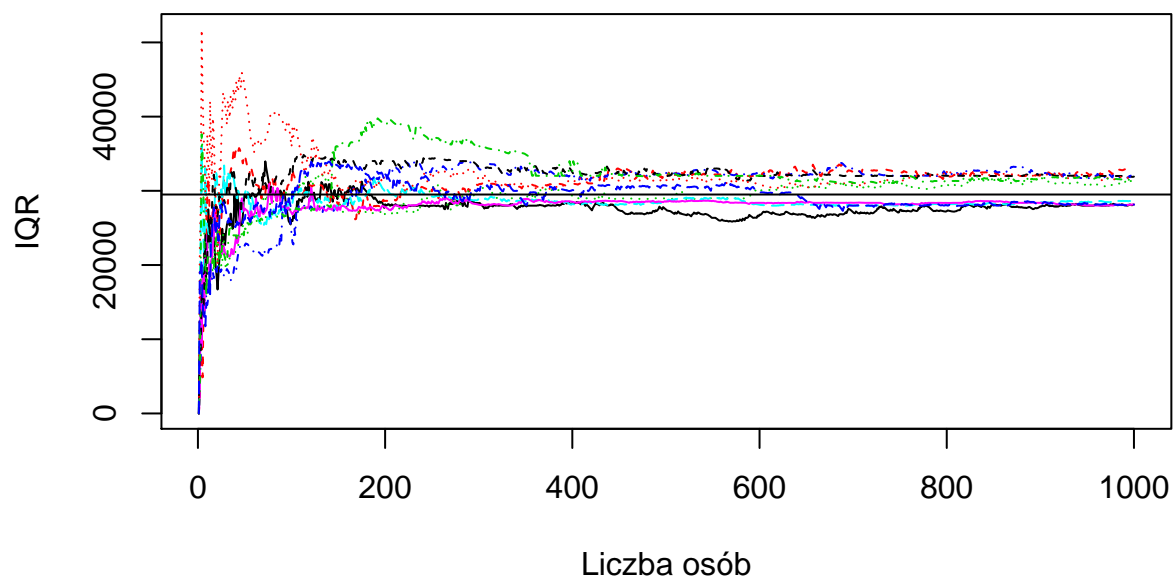
Trajektorie redniej



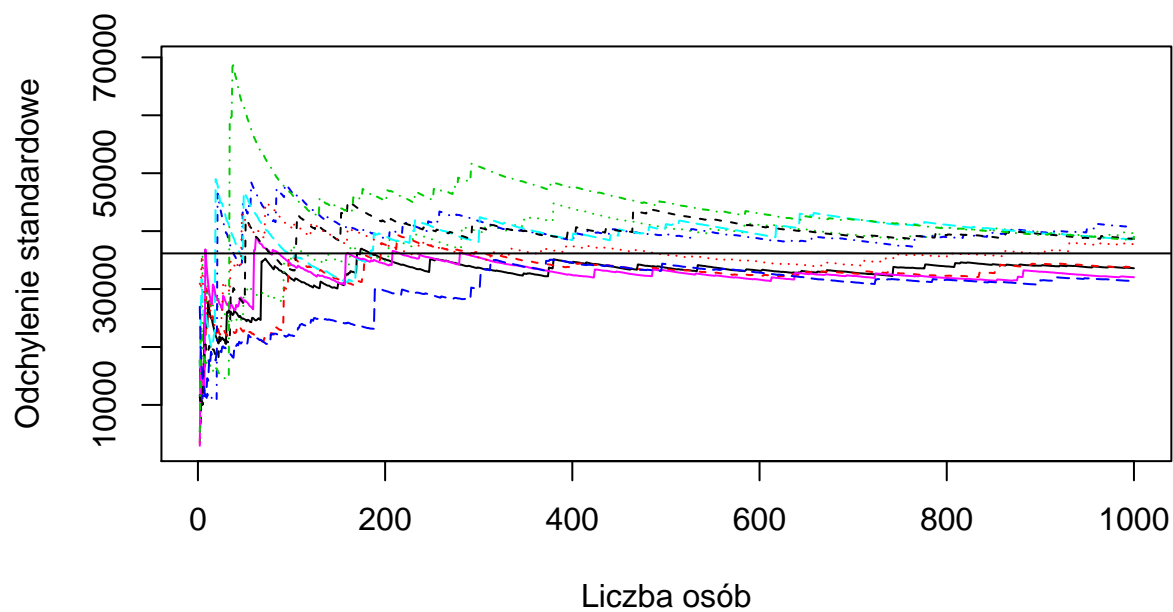
Trajektorie mediany



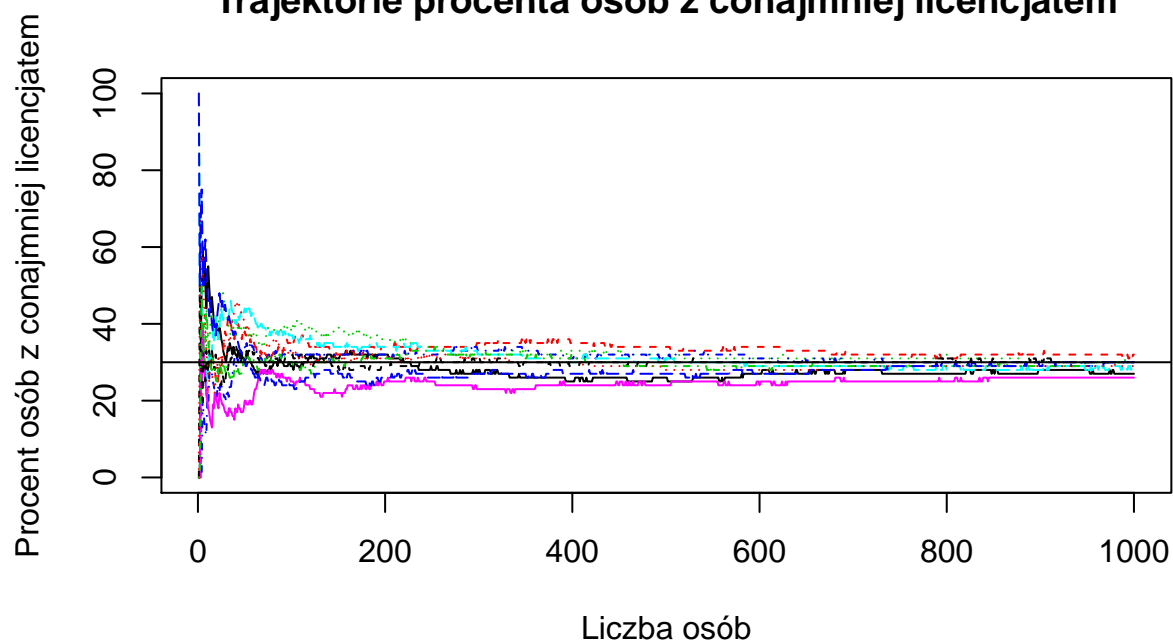
Trajektorie IQR



Trajektorie odchylenia standardowego



Trajektorie procenta osób z conajmniej licencjatem



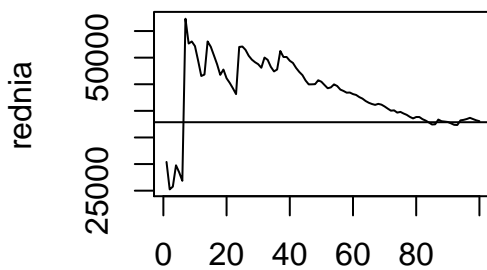
Widać, że wraz ze wzrostem rozmiaru próby odpowiednie statystyki zbiegają do statystyk wyliczonych na całym zbiorze.

b)

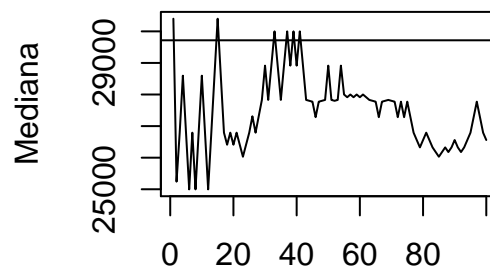
```
p = round(table(tab[[6]])/dim(tab)[1] * 1000)
l = list(sample(which(tab[6] == names(p)[[1]]), p[1]),
         sample(which(tab[6] == names(p)[[2]]), p[2]),
         sample(which(tab[6] == names(p)[[3]]), p[3]))
for(i in 1:10) {
  p = round(table(tab[, 6])/dim(tab)[1] * i*100)
  indices = sample(c(1[[1]][1:p[1]], 1[[2]][1:p[2]], 1[[3]][1:p[3]]))
  s = tab[indices,]
  v = sapply(1:(i*100), function(x) c(mean(s[1:x,5]),
                                       median(s[1:x,5]), IQR(s[1:x,5]), sd(s[1:x,5]),
                                       at_least_bachelor(s[1:x,])))

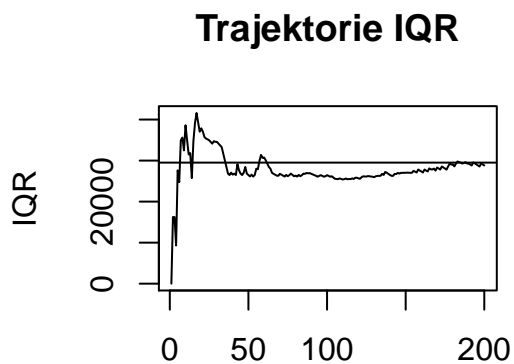
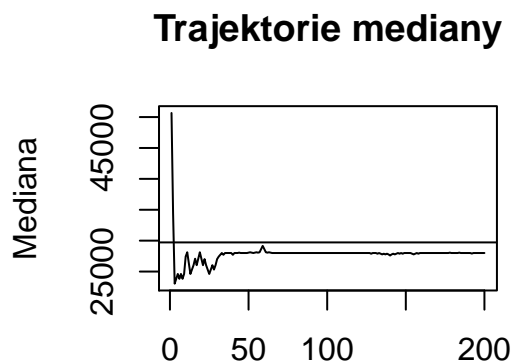
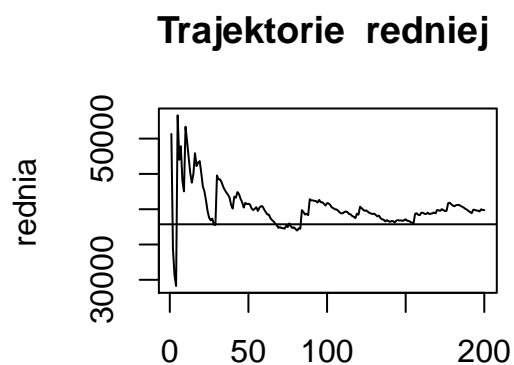
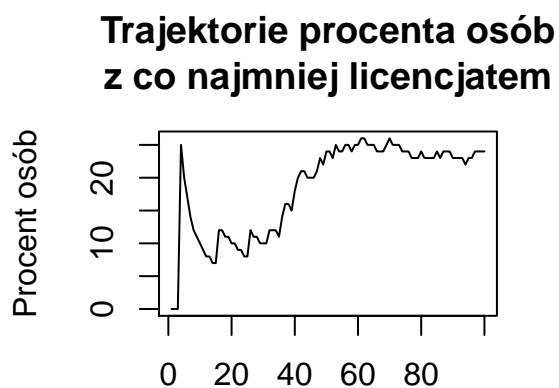
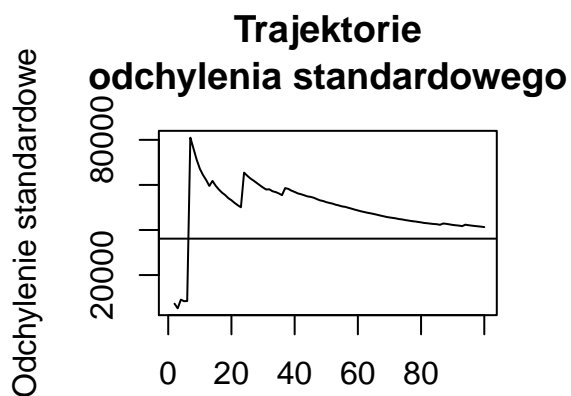
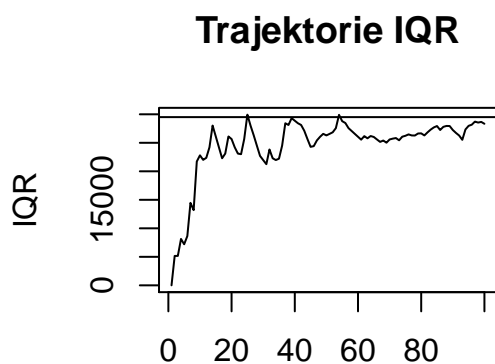
  matplot(v[1,], type = "l", main = "Trajektorie średniej", ylab = "Średnia")
  abline(income_mean, 0)
  matplot(v[2,], type = "l", main = "Trajektorie mediany", ylab = "Mediana")
  abline(income_median, 0)
  matplot(v[3,], type = "l", main = "Trajektorie IQR", ylab = "IQR")
  abline(income_iqr, 0)
  matplot(v[4,], type = "l", main = "Trajektorie\ndochylenia standardowego",
          ylab = "Odchylenie standardowe")
  abline(income_sd, 0)
  matplot(v[5,], type = "l", main = "Trajektorie procenta osób\ńz co najmniej licencjatem",
          ylab = "Procent osób")
  abline(at_least_bachelor_percent, 0)
}
```

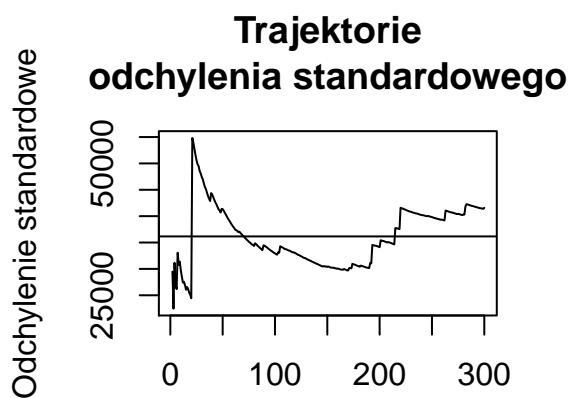
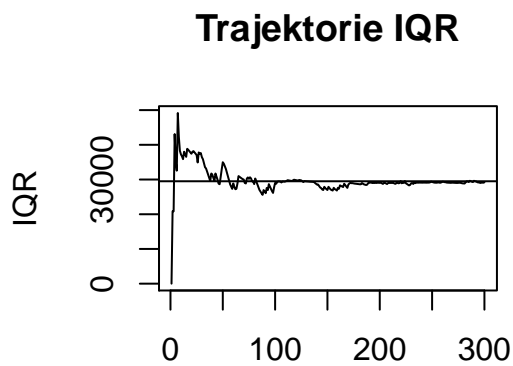
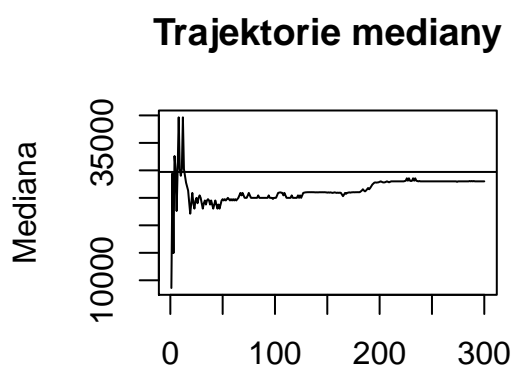
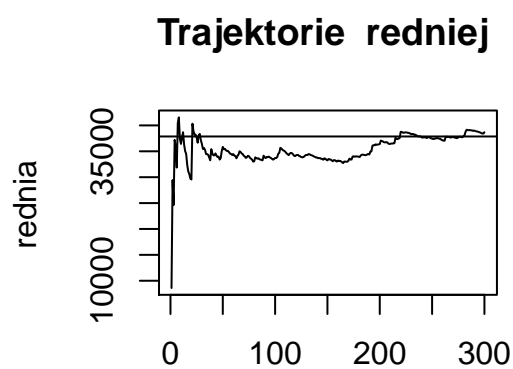
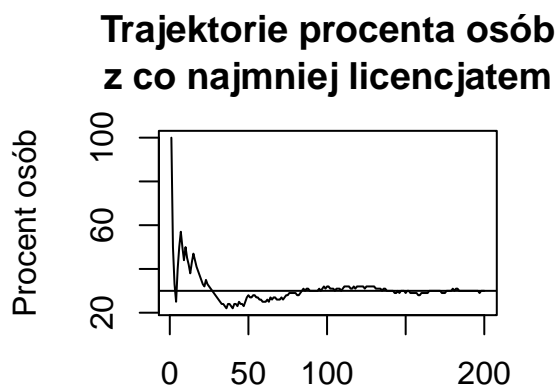
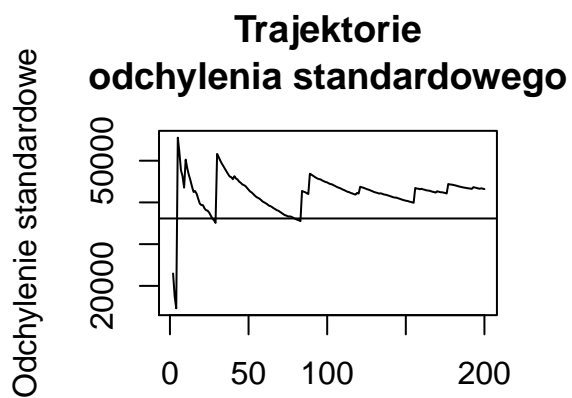
Trajektorie redniej



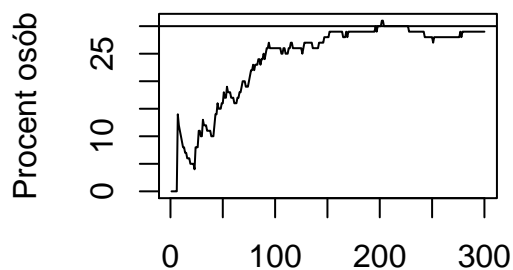
Trajektorie mediany



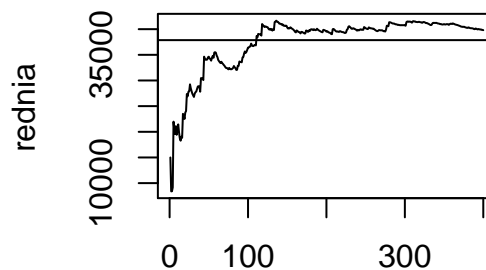




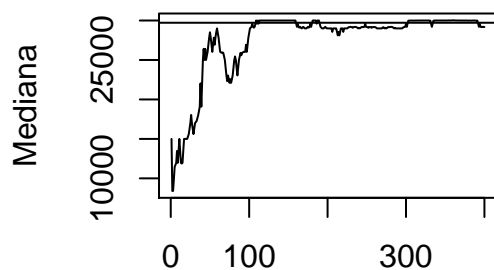
**Trajektorie procenta osób
z co najmniej licencjatem**



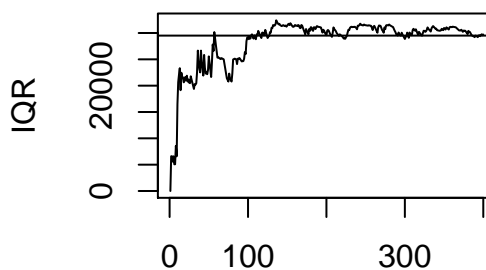
Trajektorie redniej



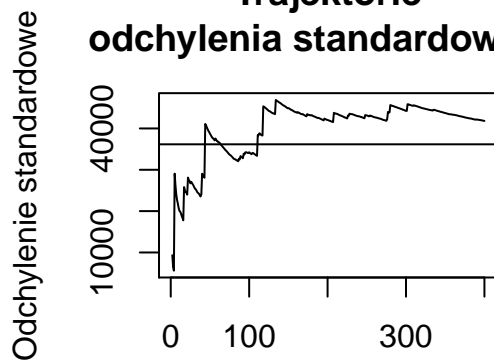
Trajektorie mediany



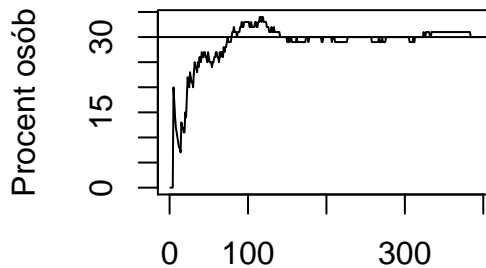
Trajektorie IQR



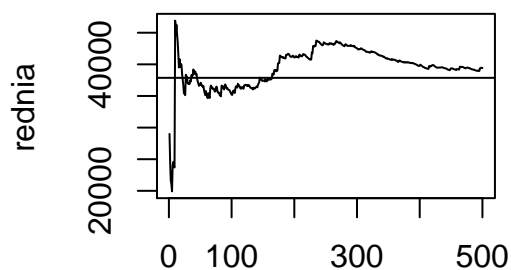
**Trajektorie
odchylenia standardowego**



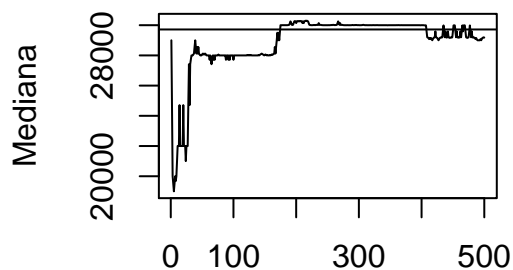
**Trajektorie procenta osób
z co najmniej licencjatem**



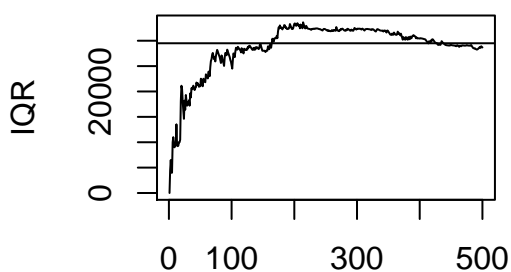
Trajektorie redniej



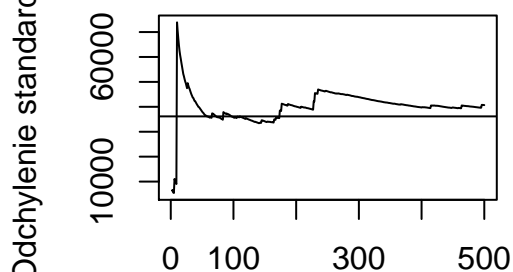
Trajektorie mediany



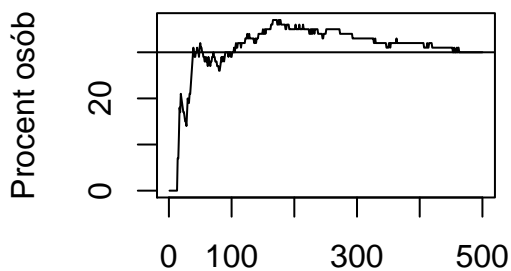
Trajektorie IQR



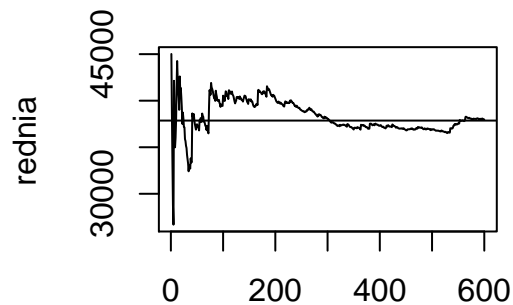
Trajektorie odchylenia standardowego



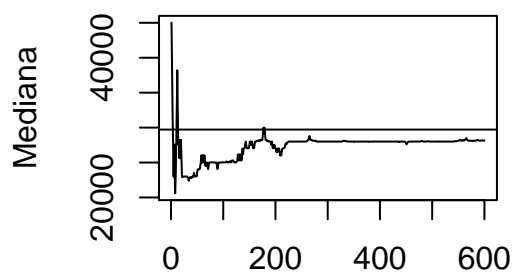
Trajektorie procenta osób z co najmniej licencjatem



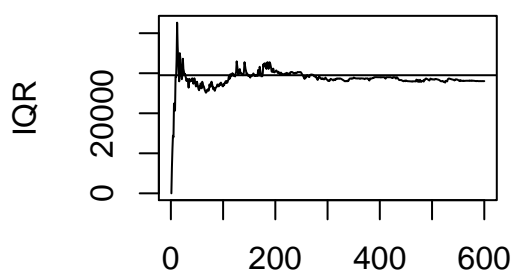
Trajektorie redniej



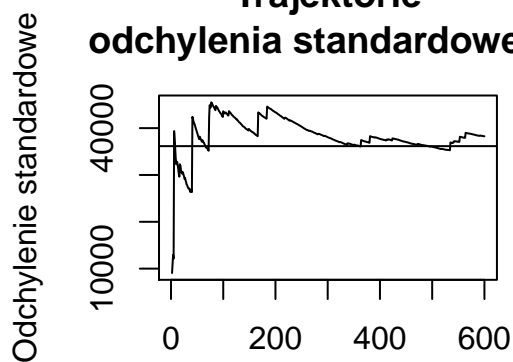
Trajektorie mediany



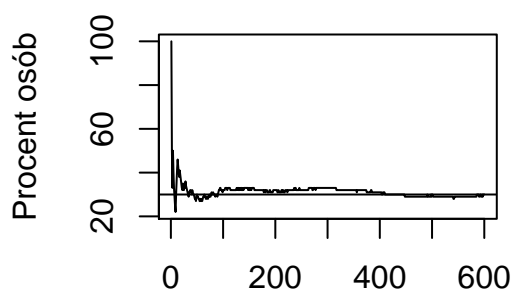
Trajektorie IQR



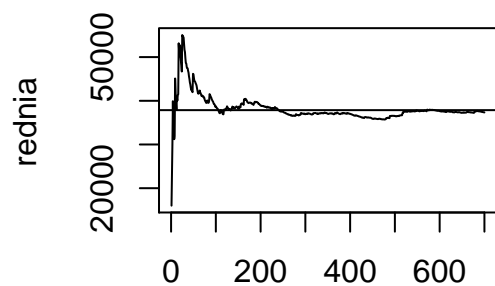
Trajektorie odchylenia standardowego



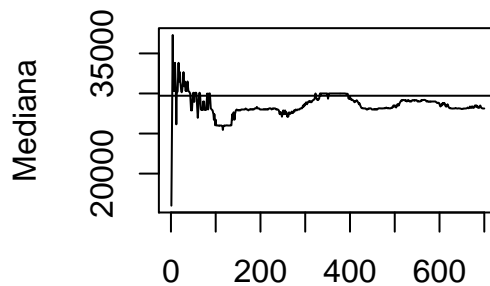
Trajektorie procenta osób z co najmniej licencjatem

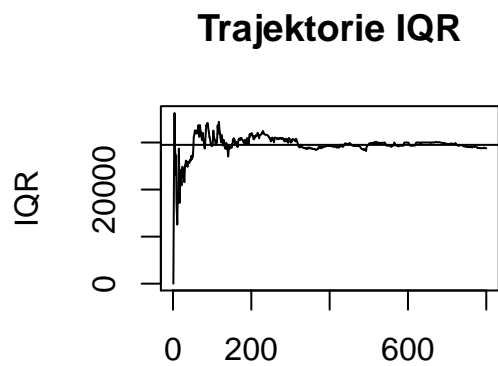
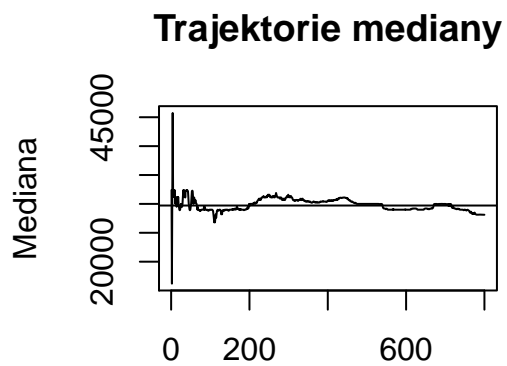
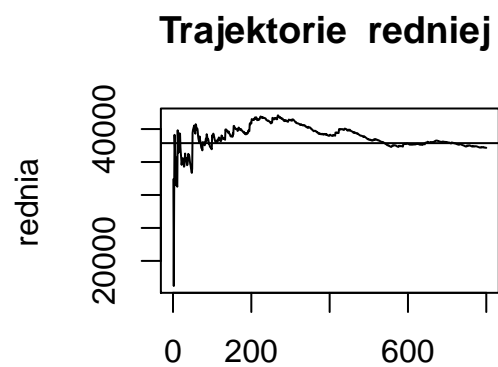
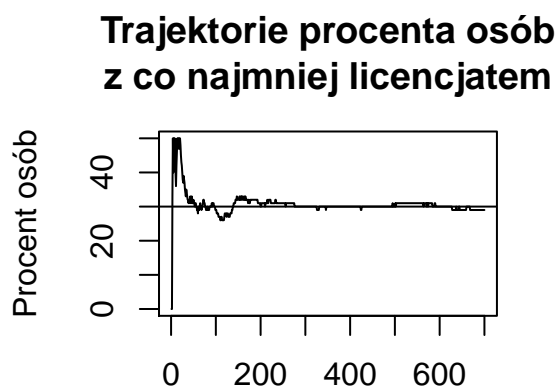
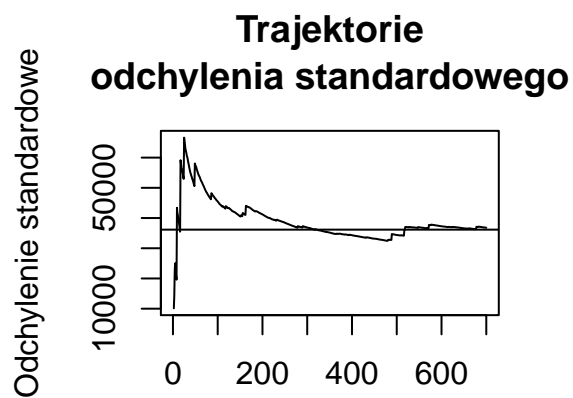
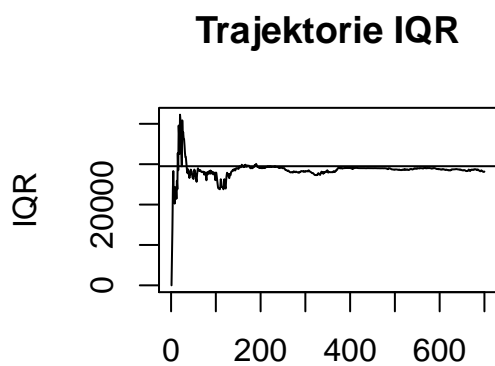


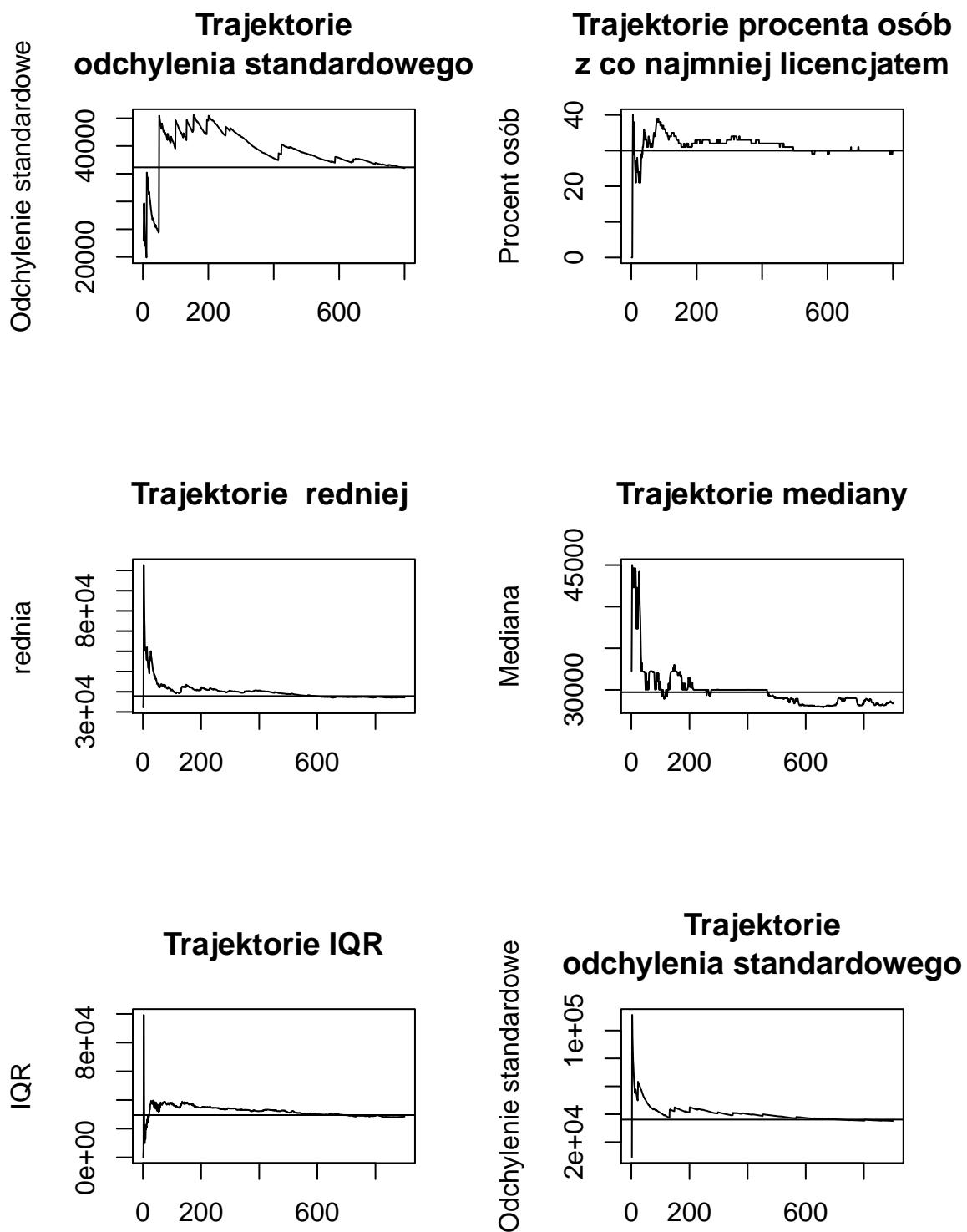
Trajektorie redniej

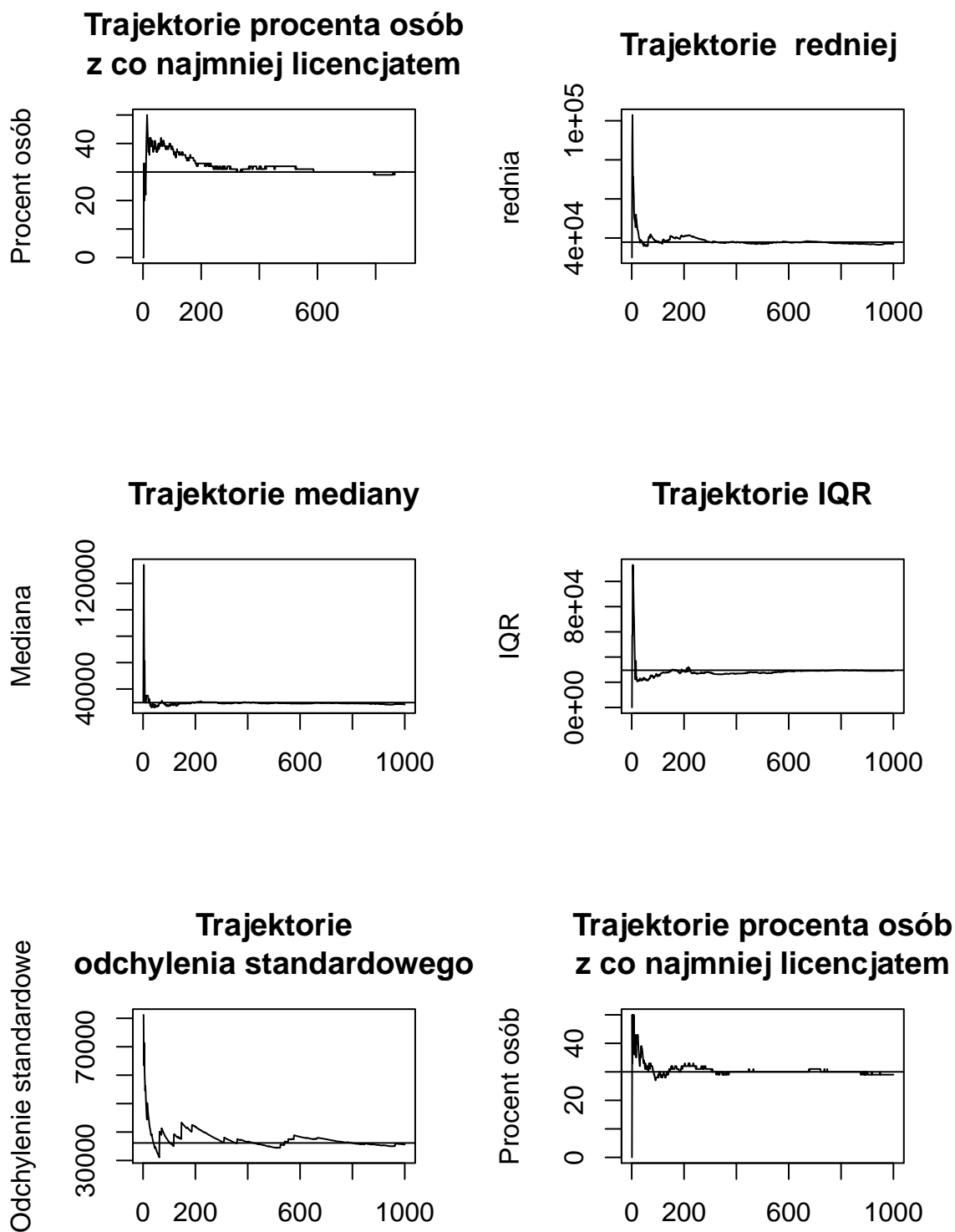


Trajektorie mediany







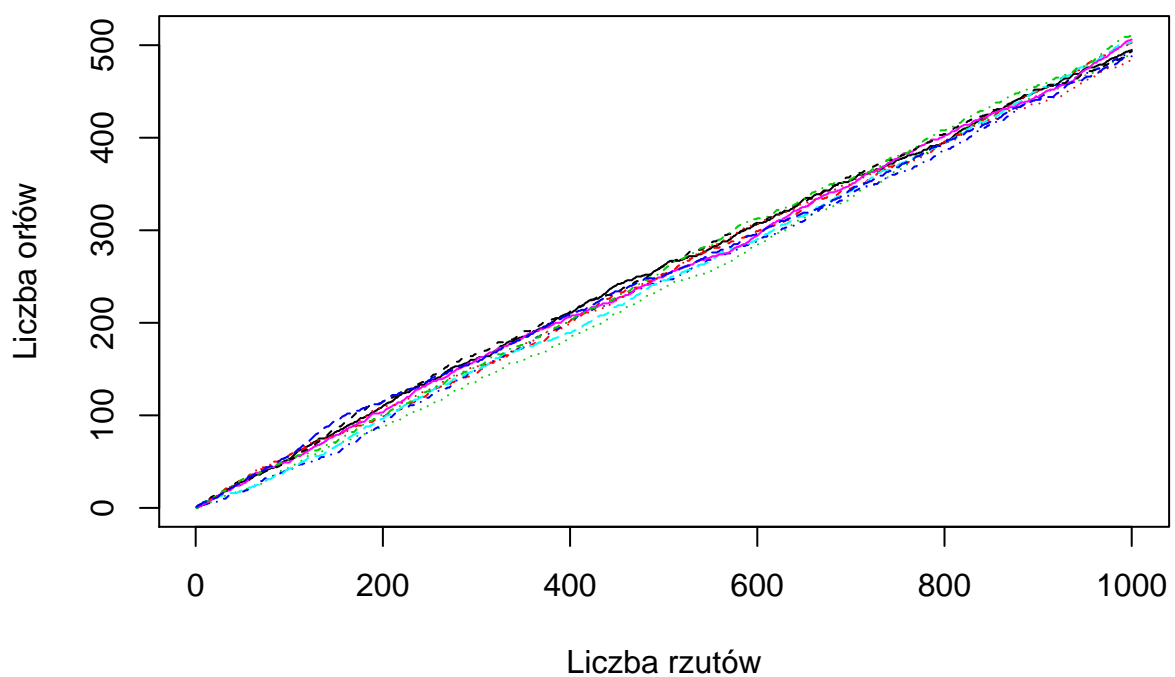


Podobnie jak w podpunkcie a statystyki z poszczególnych prób wraz ze wzrostem liczby reprezentantów zbiegają do wartości odpowiednich statystyk w całym zbiorze.

Zadanie 2

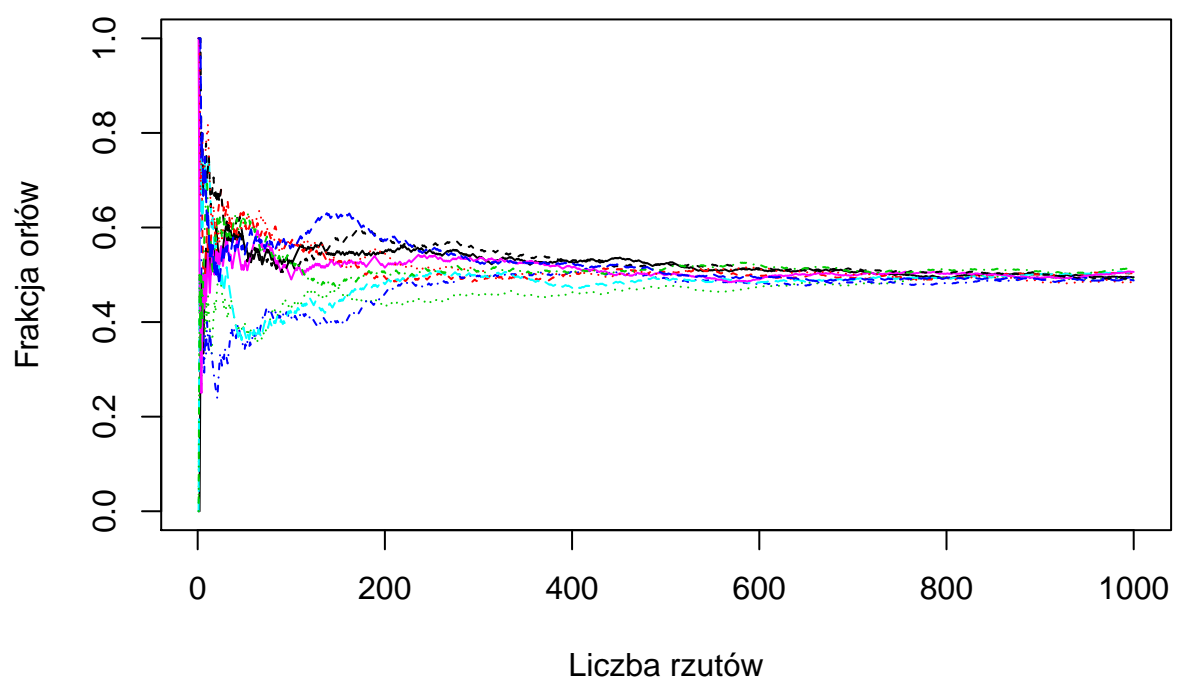
```
u = array(0, c(10, 2, 1000))
for(i in 1:10)
{
  s = sample(0:1, 10000, replace = TRUE)
  u[i,,] = sapply(1:1000, function(i) c(sum(s[1:i]), mean(s[1:i])))
}
matplot(t(u[,1,]), type = "l", main = "Trajektorie liczby orłów",
        ylab = "Liczba orłów", xlab = "Liczba rzutów")
```

Trajektorie liczby orłów



```
matplot(t(u[,2,]), type = "l", main = "Trajektorie frakcji orłów",
        ylab = "Frakcja orłów", xlab = "Liczba rzutów")
```

Trajektorie frakcji orłów



Widzimy, że frakcja liczby orłów zbiega do 0.5 czyli wartości oczekiwanej liczby orłów. Widzimy też, że liczba wystąpień orłów wzrasta wprost proporcjonalnie do liczby wykonanych rzutów.