

Vaalinumerot

Jakke Mäkelä

2021-05-10

Lasketaan tietyn vaalinumeron saamisen todennäköisyys, kun henkilön paikka omalla listalla tiedetään, ja listojen henkilömäärät tiedetään.

```
#Nopeutetaan laskentaa koodaamalla saman pituiset listat samalla koodilla. Käytetään vektoreita nopeuttamaan.
#Huom! Oletetaan, että oma lista on koodilla numero 1
listaKoodi <- c(1, 1, 1, 1, 1, 2, 3, 4, 5, 6, 6, 7, 8,8)
listaPituus<- c(100, 71, 58, 21, 8, 4, 2, 1)

omaKoodi <- 1 #Listan koodi jolla itse on
omaSijoitus <- 56 #Oma sijainti omalla listalla

pieninNumeroKorjaus <- 1 #Ykkösen ja pienimmän jaetun ehdokasnumeron ero. Normaalisti pienin ehdokasnumero

listaLukumaara <- length(listaKoodi) #nopeuttamiseksi tehdään oma muuttuja

ehdokkaidenMaara <- sum(listaPituus[listaKoodi])
uniikitListaPituudet <- length(unique(listaKoodi))

uniikkienKombinaatioidenMaara <- factorial(uniikitListaPituudet)

omaNumeroSimuloitu <- NULL

for (kokeilu in 1:uniikkienKombinaatioidenMaara){

  otos <- sample(listaKoodi) #Oletuksena järjestää luvut

  #Yhdellä askeleella:
  #Lasketaan slotin otos[i] listan pituus
  #Lasketaan yhteen millä numerolla lista otos[i] alkaa, tehden alkunumeron korjaus
  #Poistetaan viimeinen cumsum turhana
  alkuKohta <- (c(1,cumsum(listaPituus[otos]))+pieninNumeroKorjaus)[1:listaLukumaara]

  #Kaikki mahdolliset omat numerot, lisätään vektoriin. Vähennetään yksi jotta sijaluku tulee oikein.
  omaNumeroSimuloitu <- c(omaNumeroSimuloitu,alkuKohta[otos==omaKoodi]+omaSijoitus-1)

}

pNumero <- as.data.frame(table(omaNumeroSimuloitu)/length(omaNumeroSimuloitu)) %>%
  mutate(omaNumero=as.numeric(as.character(omaNumeroSimuloitu))) %>%
  select(omaNumero,Freq)
```

```

todennak <- pNumero %>%
  arrange(desc(Freq)) %>%
  rename(prosentti=Freq) %>%
  mutate(prosentti=round(100*prosentti,1)) %>%
  mutate(loppuKaksi = omaNumero %% 100) %>%
  mutate(loppuYksi = omaNumero %% 10)

write.csv(todennak,file="Todennakoisyydet.csv")

loppuKaksi <- todennak %>% group_by(loppuKaksi) %>% summarize(prosentti=sum(prosentti)) %>% arrange(desc(prosentti))

loppuYksi <- todennak %>% group_by(loppuYksi) %>% summarize(prosentti=sum(prosentti)) %>% arrange(desc(prosentti))

todennak_print <- todennak %>%
  mutate(yhteensa=cumsum(prosentti)) %>%
  select(omaNumero,prosentti,yhteensa)
print(head(todennak_print,20))

```

```

##      omaNumero  prosentti yhteensa
## 1          57        8.2      8.2
## 2         626        7.2     15.4
## 3         156        2.2     17.6
## 4         526        2.2     19.8
## 5          60        1.1     20.9
## 6         622        1.1     22.0
## 7         625        1.1     23.1
## 8         160        0.8     23.9
## 9         522        0.8     24.7
## 10        157        0.7     25.4
## 11        525        0.7     26.1
## 12         58        0.7     26.8
## 13        618        0.7     27.5
## 14         64        0.7     28.2
## 15        256        0.7     28.9
## 16        426        0.7     29.6
## 17        624        0.6     30.2
## 18        555        0.6     30.8
## 19        127        0.6     31.4
## 20        114        0.6     32.0

```

```

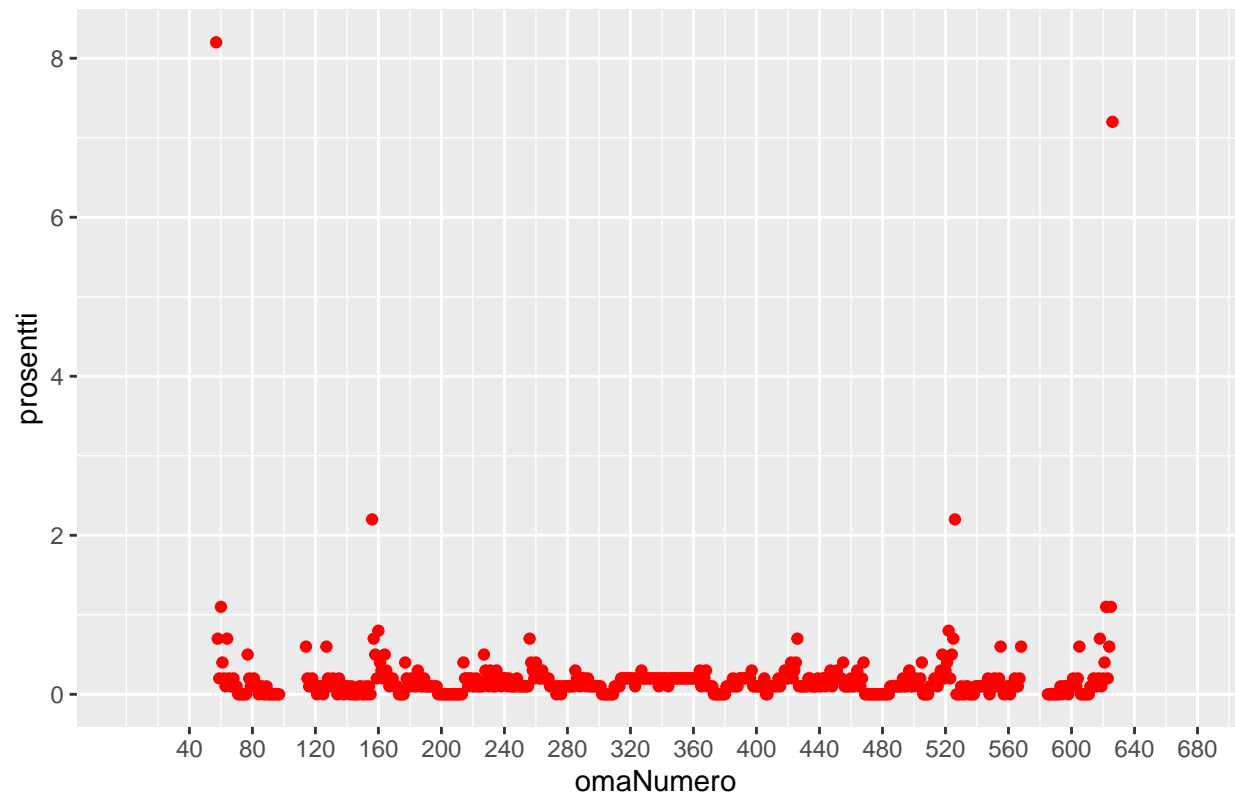
gPlot <- ggplot(todennak)+
  geom_point(aes(x=omaNumero,y=prosentti),col="red")+
  coord_cartesian(xlim=c(1+pieninNumeroKorjaus,ehdokkaidenMaara))+
  scale_x_continuous(breaks=seq(40,700,40))+

  ggtitle("Numeron todennäköisyys")

print(gPlot)

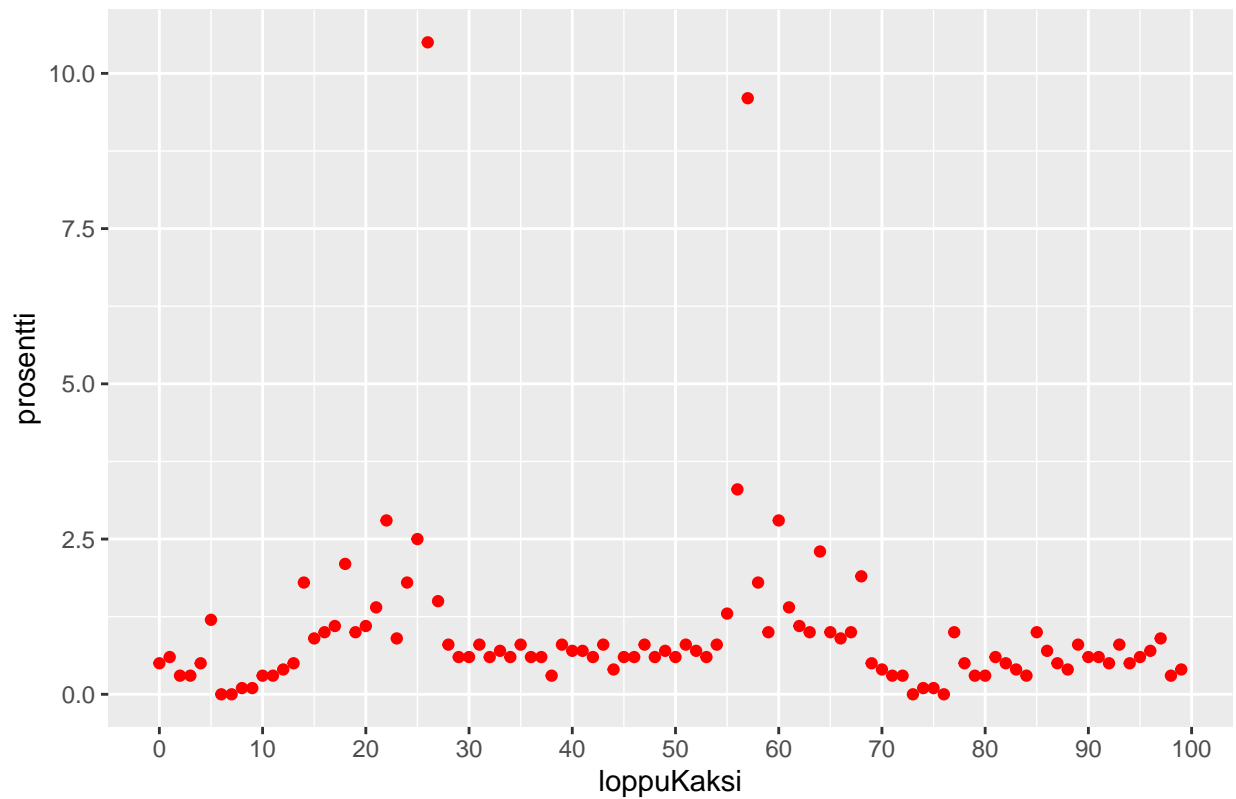
```

Numeron todennäköisyys



```
gPlot2 <- ggplot(loppuKaksi)+  
  geom_point(aes(x=loppuKaksi,y=prosentti),col="red")+  
  scale_x_continuous(breaks=seq(0,100,10))+  
  
  ggtitle("Kahden viimeisen numeron todennäköisyys")  
print(gPlot2)
```

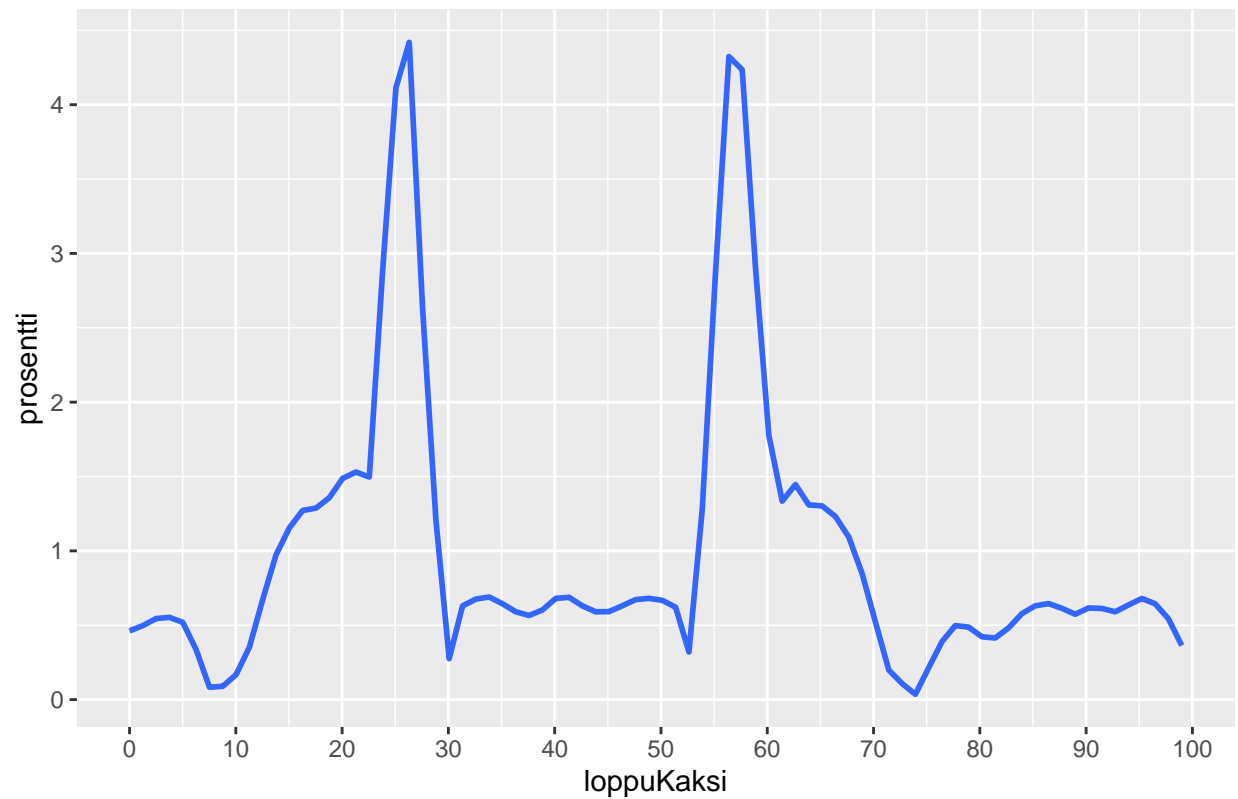
Kahden viimeisen numeron todennäköisyys



```
gPlot3 <- ggplot(loppuKaksi)+  
  geom_smooth(aes(x=loppuKaksi,y=prosentti),span=0.1,se=FALSE)+  
  scale_x_continuous(breaks=seq(0,100,10))+  
  ggtitle("Kahden viimeisen numeron todennäköisyys")  
  
print(gPlot3)
```

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

Kahden viimeisen numeron todennäköisyys



```
gPlot4 <- ggplot(loppuYksi)+  
  geom_point(aes(x=loppuYksi,y=prosentti),col="red")+  
  scale_x_continuous(breaks=seq(0,10,1))+  
  
  ggtitle("Viimeisen numeron todennäköisyys")  
  
print(gPlot4)
```

Viimeisen numeron todennäköisyys

