R final

최수민(201718133)

전북대학교 사회학과

12/21/2021

```r
#install.packages("dplyr")
#install.packages("tidyverse")
library(dplyr)
```

```
## Warning: 패키지 'dplyr'는 R 버전 4.1.2에서 작성되었습니다
```

```
##
## 다음의 패키지를 부착합니다: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyverse)
```

```
## Warning: 패키지 'tidyverse'는 R 버전 4.1.2에서 작성되었
습니다
```

```
## -- Attaching packages ---------------------------------
```

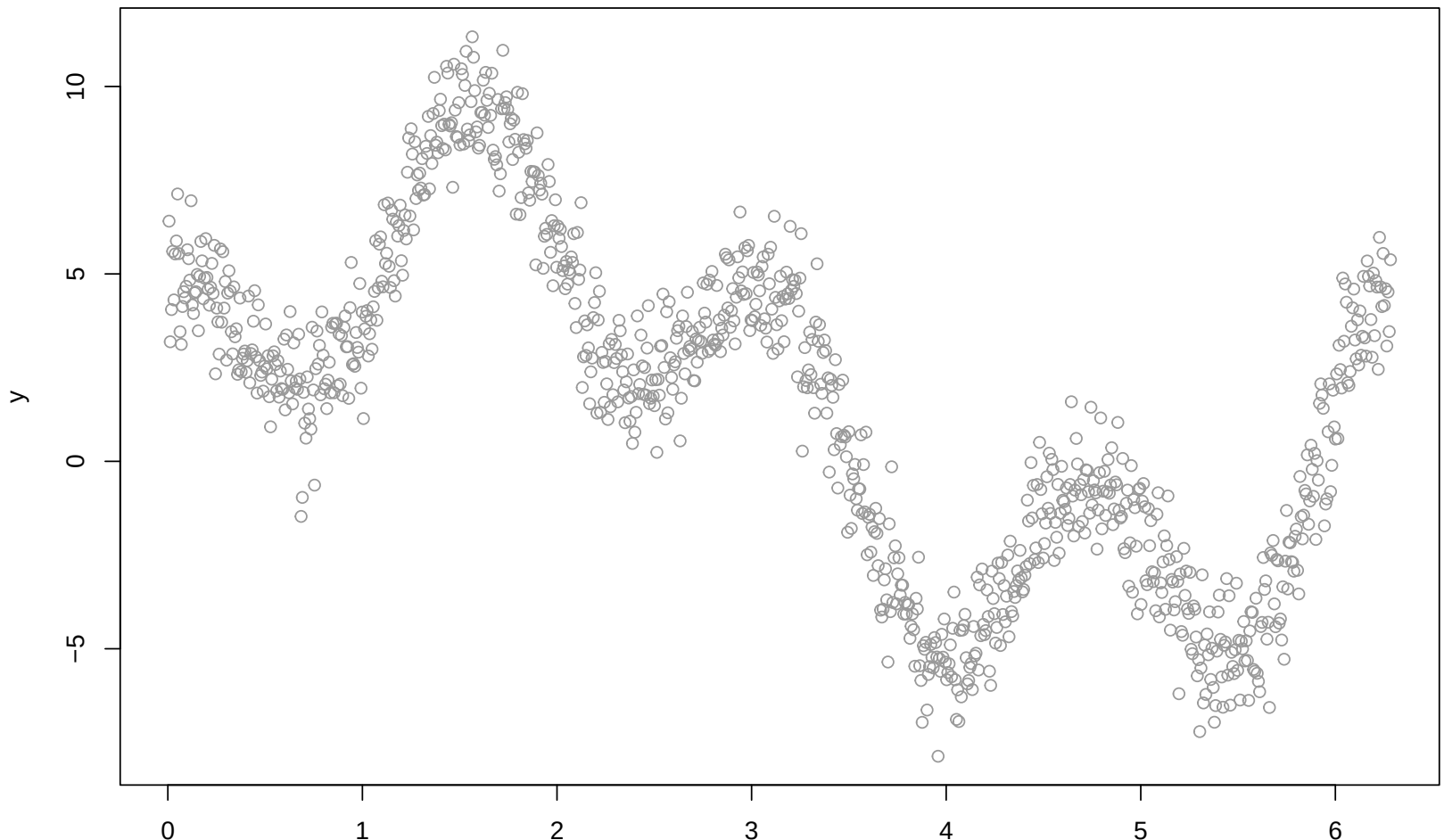
1번

```
y = c()
t = c()
for (i in 1:1000){
  t[i] = (2*pi*i)/1000
  y[i] = 1.5+(5*sin(t[i]))+(3*cos(4*t[i]))
}
```

```
epsilon = rnorm(1000,mean=0,sd=1)
y = y+epsilon
```

```
x_1 = c()
x_2 = c()
for (i in 1:1000){
  x_1[i] = sin(t[i])
  x_2[i] = cos(4*t[i])
}
```

```
for (i in 1:1000){
  y[i] = 1.5+(5*x_1[i])+(3*x_2[i])
}
y = y+epsilon
plot(t,y,col='gray60')
```

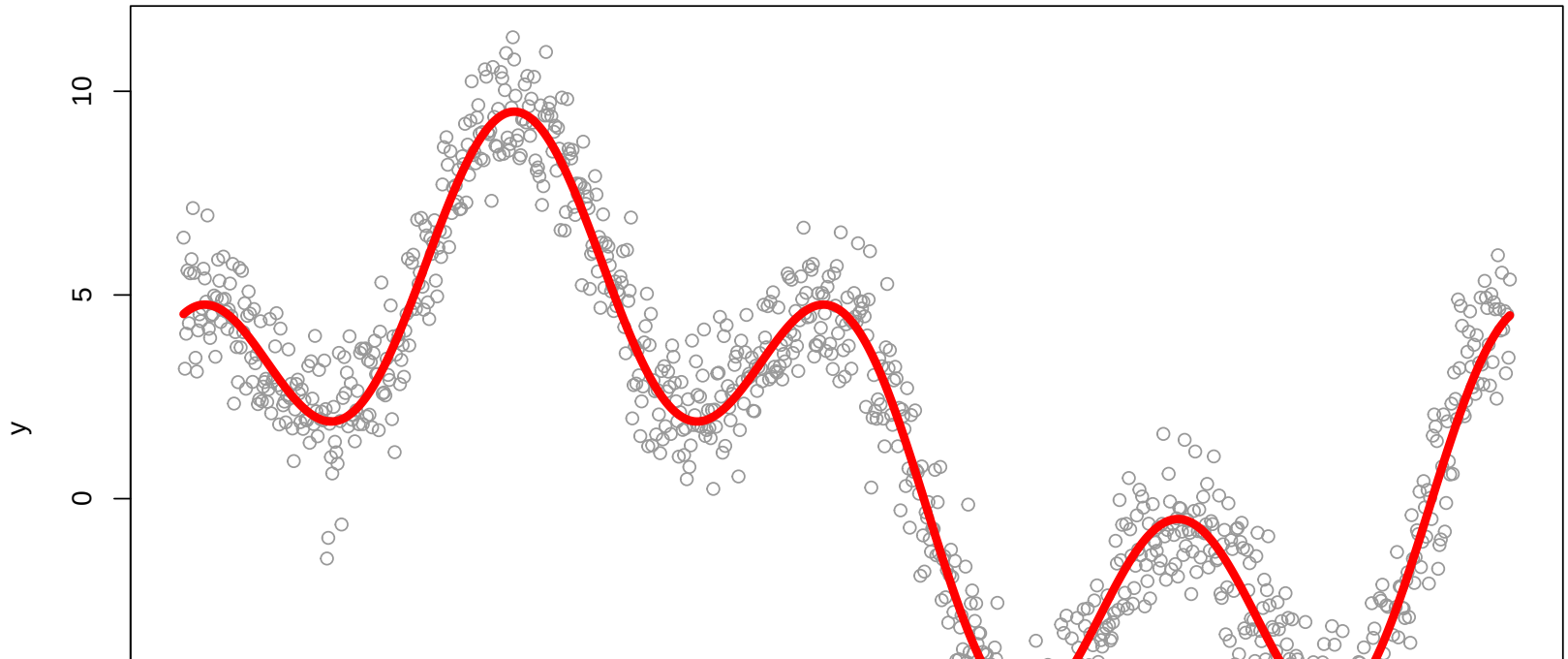


```r
X = cbind(1,x_1,x_2)
head(X)
```

```
##                   x_1       x_2
## [1,] 1 0.006283144 0.9996842
## [2,] 1 0.012566040 0.9987370
## [3,] 1 0.018848440 0.9971589
## [4,] 1 0.025130095 0.9949510
## [5,] 1 0.031410759 0.9921147
## [6,] 1 0.037690183 0.9886517
```

```
beta = rbind(1.5,5,3)
rslt = X %*% beta

for (i in 1:1000){
  y[i] = 1.5+(5*x_1[i])+(3*x_2[i])
}
y = y+epsilon
plot(t,y,col='gray60')
lines(t,rslt,col='red',lwd=5)
```



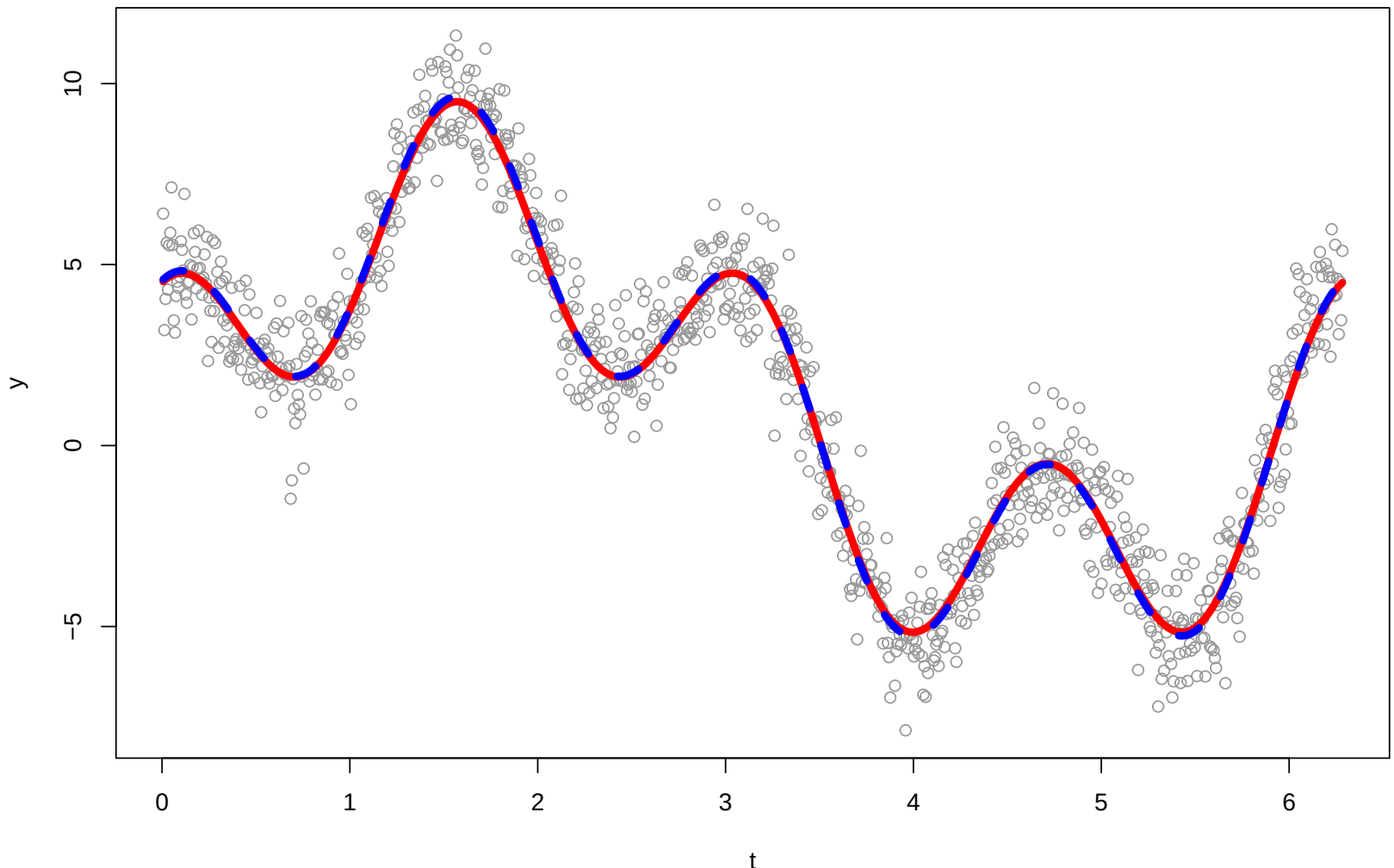
```
betahat = solve(t(X) %*% X) %*% t(X) %*% y
print(beta)
```

```
##      [,1]
## [1,]  1.5
## [2,]  5.0
## [3,]  3.0
```

```
print(betahat)
```

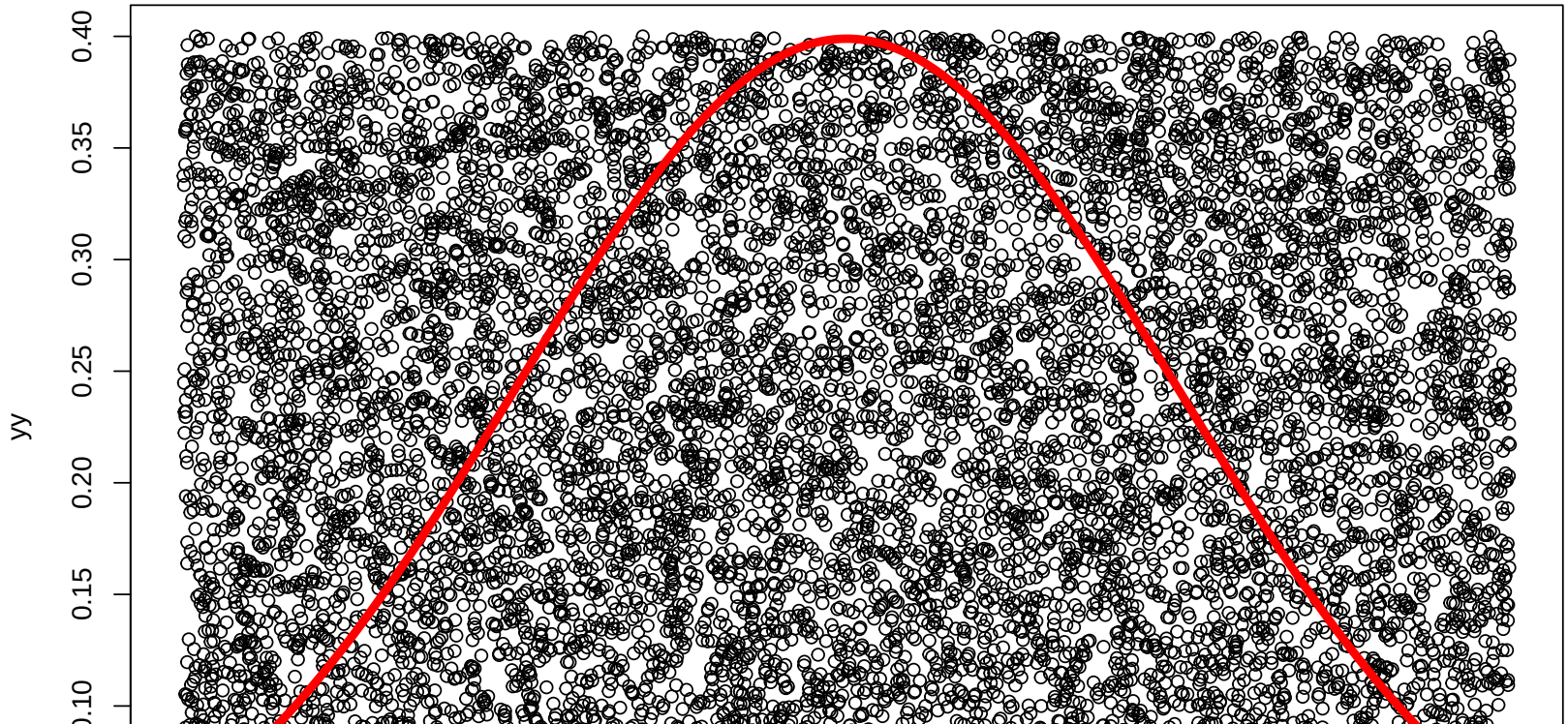
```
##          [,1]
##     1.509969
## x_1 5.082760
## x_2 3.049195
```

```
rslt_ = X %*% betahat
plot(t,y,col='gray60')
lines(t,rslt,col='red',lwd=5)
lines(t,rslt_,col='blue',lty=2,lwd=5)
```

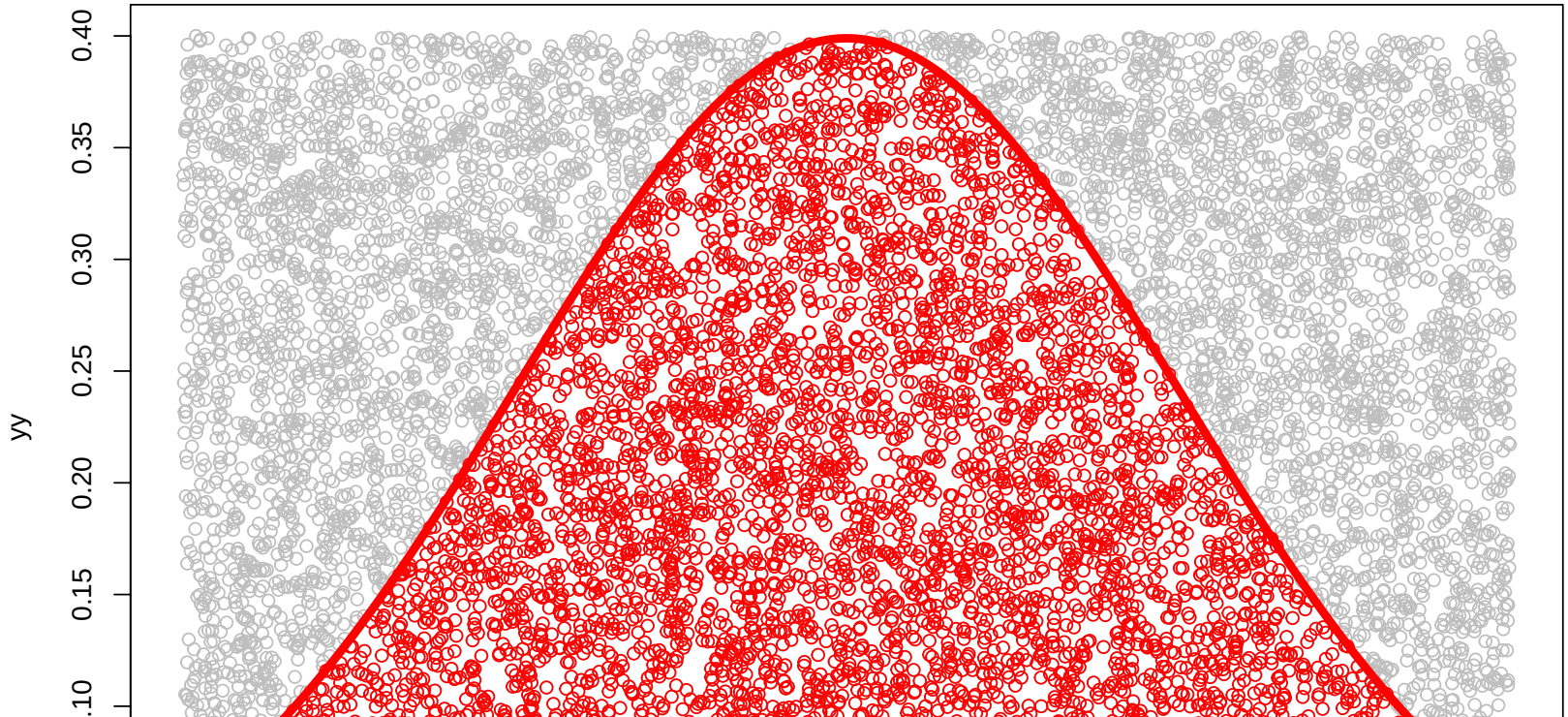


2번

```
x=seq(from=-1.96,to=1.96,by=0.01)
y=(1/sqrt(2*pi))*exp(-1/2*x^2)
xx = runif(10000, -2, 2)
yy = runif(10000, 0.05, 0.4)
plot(xx,yy)
lines(x,y,col='red',lwd=5)
```



```
test = function(xx,yy) {
  yy < (1/sqrt(2*pi))*exp(-1/2*xx^2)
}
tst = c()
for (i in 1:10000) tst[i] = test(xx[i],yy[i])
plot(xx,yy,col='gray')
lines(x,y,col='red',lwd=5)
points(xx[tst],yy[tst],col='red')
```



```r
print(sum(tst))
```

```
## [1] 5300
```

```r
sum(tst)/10000 * 4 * 0.35 #넓이
```

```
## [1] 0.742
```

```r
a = rnorm(1000,mean=0,sd=1)
head(a)
```

```
## [1] -0.5704097 -0.3201172  1.5479921 -0.5991075 -0.4610031  1
```

```r
aa = (a[a > -1.96 & a < 1.96])
length(aa)
```

```
## [1] 947
```

3번

```
#type A
ARR = c('N1','N2','N3','N4','N5','N6','N7','N8','A','N9')
SURV = 10
PLAYER = ARR[SURV]
STAGE = 0
tossRSLT = NA

reset = function(){
SURV <<- 10
STAGE <<- 0
PLAYER <<- ARR[SURV]
tossRSLT <<- NA
}
toss = function(p) rbinom(n=1, size=1, prob=p) %>% as.logical
```

```
go = function(){
  for (i in 1:20){
    PROB = 0.5+(PLAYER=='A')*0.45
    tossRSLT <<- toss(PROB)
    if (tossRSLT == FALSE) SURV <<- SURV-1
    STAGE <<- STAGE + 1
    PLAYER <<- ARR[SURV]
    if (SURV == 0) break
  }
}
simulate_once = function(){
  reset()
  go()
  return(record()$SURV)
}
record = function(){
  list(SURV=SURV, STAGE=STAGE)
  }

surv8 = 0
```

```r
for (i in 1:10000){
  reset()
  go()
  record()
  if (record()$SURV > 7) surv8 <<- surv8 + 1
}

print(surv8/10000)
```

```
## [1] 0.4376
```

```r
#type B
ARR = c('N9','A','N8','N7','N6','N5','N4','N3','N2','N1')
SURV <<- 10
PLAYER <<- ARR[SURV]
STAGE <<- 0
tossRSLT <<- NA
surv8 <<- 0

for (i in 1:10000){
  reset()
  go()
  record()
  if (record()$SURV <= 3) surv8 <<- surv8 + 1
}

print(surv8/10000)
```

```
## [1] 0.9435
```

4번

```
df=read_csv('https://raw.githubusercontent.com/guebin/2021IR/mas
```

```
## Rows: 12294 Columns: 5
```

```
## -- Column specification --------------------------------
## Delimiter: ","
## chr (1): prov
## dbl (4): year, month, day, cases
```

```
##
## i Use `spec()` to retrieve the full column specification for
## i Specify the column types or set `show_col_types = FALSE` to
```

```
head(df)
```

```
## # A tibble: 6 x 5
##    year month   day prov  cases
##   <dbl> <dbl> <dbl> <chr> <dbl>
## 1  2020     1    20 서울      0
## 2  2020     1    20 부산      0
## 3  2020     1    20 대구      0
```

```
df %>% group_by(year) %>% summarize(sum=sum(cases))
```

```
## # A tibble: 2 x 2
##    year    sum
##   <dbl>  <dbl>
## 1  2020  60726
## 2  2021 396886
```

```
df %>% group_by(prov) %>% filter(year==2020, month==2, day<=15)
```

```
## # A tibble: 18 x 2
##    prov    sum
##    <chr> <dbl>
##  1 강원      0
##  2 검역      0
##  3 경기      9
##  4 경남      0
##  5 경북      0
##  6 광주      2
##  7 대구      0
##  8 대전      0
##  9 부산      0
## 10 서울      5
## 11 세종      0
## 12 울산      0
## 13 인천      0
## 14 전남      1
## 15 전북      0
## 16 제주      0
## 17 충남      0
```

```
df %>% group_by(prov) %>% filter(year==2020, month==2, day>15) %
```

```
## # A tibble: 18 x 2
##    prov    sum
##    <chr> <dbl>
##  1 강원      7
##  2 검역      0
##  3 경기     65
##  4 경남     59
##  5 경북    472
##  6 광주      7
##  7 대구   2055
##  8 대전     13
##  9 부산     75
## 10 서울     62
## 11 세종      1
## 12 울산     17
## 13 인천      5
## 14 전남      1
## 15 전북      4
## 16 제주      2
## 17 충남     48
```