

R입문 기말고사

허윤정(201922051)

12/20/2021

library

```
library(tidyverse)
```

1번

(1)

```
epsilon=rnorm(1000)  
head(epsilon)
```

```
## [1] -0.2292503  0.2224717  0.2326248  1.1919175 -0.2987808 -0.4737593
```

(2)

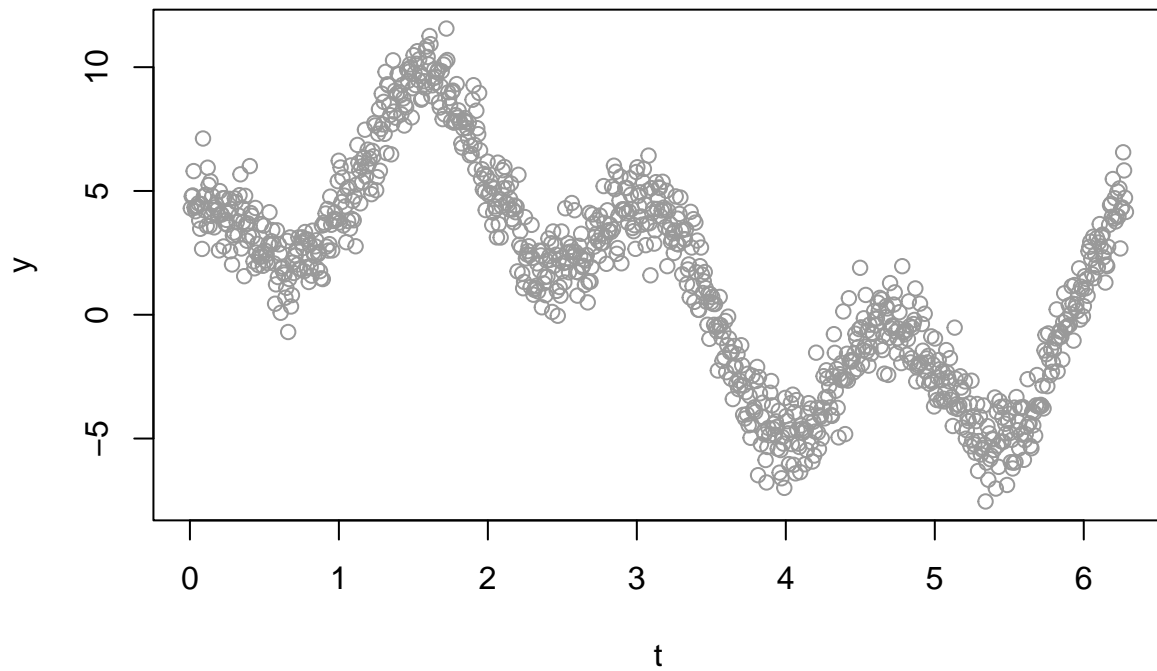
```
t=c()  
for(i in 1:1000) t[i] = 2*pi*i/1000  
  
x1=c()  
for(i in 1:1000) x1[i] = sin(t[i])  
  
x2=c()  
for(i in 1:1000) x2[i] = cos(4*t[i])  
  
head(x1);head(x2)
```

```
## [1] 0.006283144 0.012566040 0.018848440 0.025130095 0.031410759 0.037690183
```

```
## [1] 0.9996842 0.9987370 0.9971589 0.9949510 0.9921147 0.9886517
```

(3)

```
y=1.5+5*x1+3*x2+epsilon  
plot(t,y,col='gray60')
```



(4)

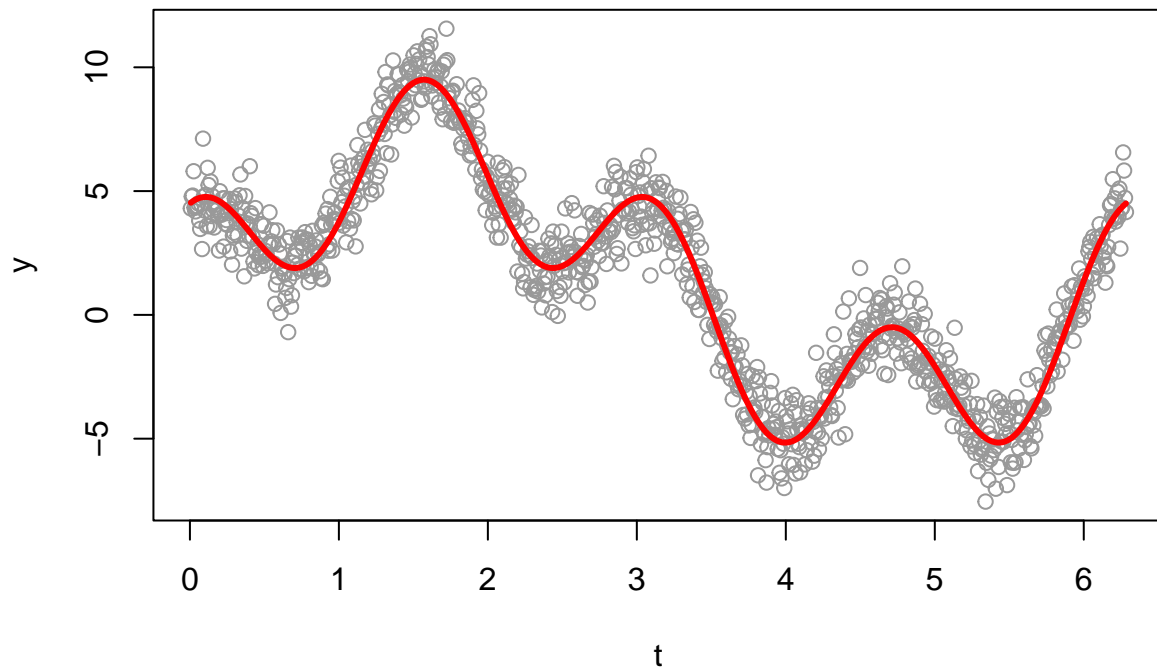
```
X=cbind(1,x1,x2)  
head(X)
```

```
##           x1           x2  
## [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
```

(5)

```
Beta=rbind(1.5,5,3)
Xbeta=c(X %*% Beta)

plot(t,y,col='gray60')
lines(t,Xbeta,col='red',lwd=3)
```



(6)

```
y=cbind(y)
hatBeta=solve(t(X)%*%X)%*%t(X)%*%y
hatBeta
```

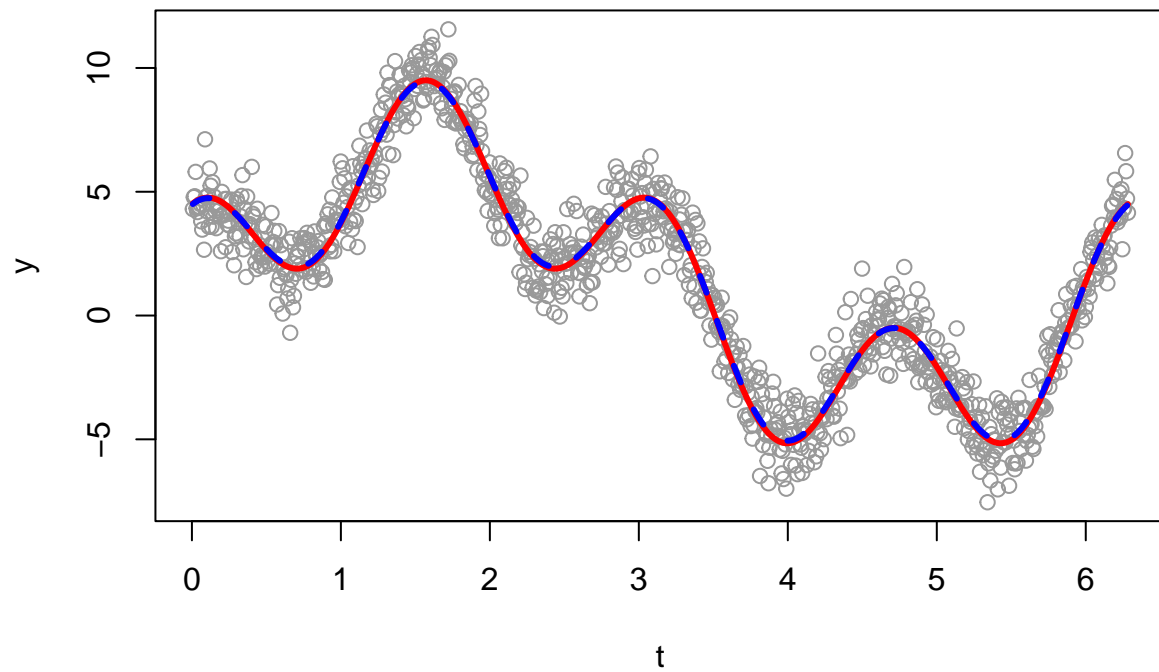
```
##          y
## 1.531120
## x1 4.979978
## x2 2.941735
```

$\hat{\beta}$ 를 β 와 비교했을때 차이가 거의 없는 것을 볼 수 있다.

(7)

```
XhatBeta=c(X %*% hatBeta)

plot(t,y,col='gray60')
lines(t,Xbeta,col='red',lwd=3)
lines(t,XhatBeta,col='blue',lty=2,lwd=3)
```

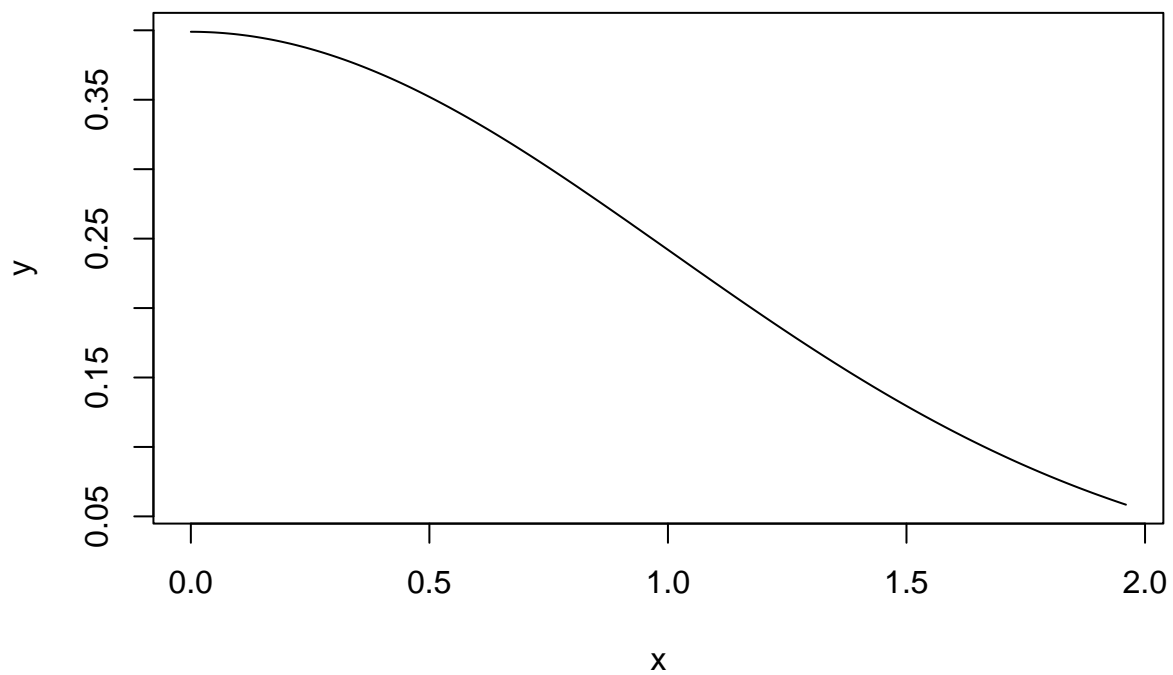


2번

(1)

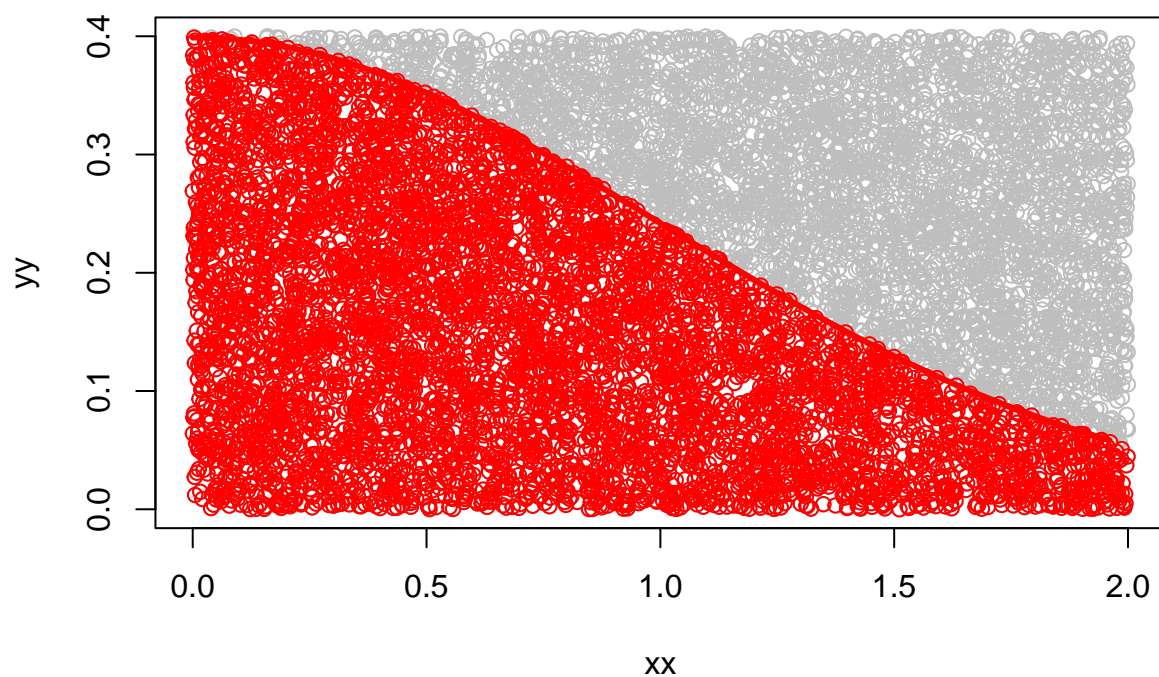
$\int_{-1.96}^{1.96} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 2 \int_0^{1.96} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$ 를 이용하여 구하였다.

```
x=seq(from=0,to=1.96,by=0.01)
y=1/sqrt(2*pi)*exp(-(x^2)/2)
plot(x,y,type='l')
```



```
xx=runif(10000)*2
yy=runif(10000)*0.4
test = function(xx,yy){
  yy < 1/sqrt(2*pi)*exp(-(xx^2)/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=3)
points(xx[tst],yy[tst],col='red')
```



```
2*(sum(tst)/10000*(2*0.4))
```

```
## [1] 0.95296
```

(2)

```
x=rnorm(1000)
sum((-1.96< x & x < 1.96)==TRUE)
```

```
## [1] 951
```

3번

```
rm(list=ls())

SIMULATE_TYPE1 = function(n){
  #변수
  ARR=c('N1','N2','N3','N4','N5','N6','N7','N8','N9','N10')
  SURV=10
  PLAYER = ARR[SURV]
  STAGE = 0
  PROB = 0.5
  TOSSRSLT = NA

  #sub_function
  toss = function(p) rbinom(n=1,size=1,prob=p) %>% as.logical
  reset = function(){
    TOSSRSLT <- NA
    SURV <- 10
    STAGE <- 0
    PLAYER <- ARR[SURV]
  }
  record = function(){
    list(PRE_TOSSRSLT=TOSSRSLT, SURV=SURV, STAGE=STAGE, PLAYER=PLAYER)
  }
  go = function(){
    PROB <- 0.5 + (PLAYER=='N9')*0.45
    TOSSRSLT <- toss(PROB)
    if (TOSSRSLT==FALSE) SURV <- SURV - 1
    STAGE <- STAGE + 1
    PLAYER <- ARR[SURV]
  }
  gogo = function() for(i in 1:10) go()

  gogo_history = function(){
    rslt_ = as_tibble(record())
    for(i in 1:10){
      go()
      rslt_ = rbind(rslt_, as_tibble(record()))
    }
    print(rslt_)
  }

  simulate_once = function(){
    reset()
  }
}
```

```

    gogo()
    return(record())$SURV )
}

#Body
simrslt = c()
for (i in 1:10000) simrslt[i] = simulate_once()
simrslt >= n #n명 이상 살아남는 경우
}

```

Type1에서 8번 참가자가 생존하려면 최소 8명 이상이 생존해야한다.

```
sum(SIMULATE_TYPE1(8))/10000
```

```
## [1] 0.7377
```


4번

```
df=read_csv('https://raw.githubusercontent.com/guebin/2021IR/master/_notebooks/covid19.c
```

```
## 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 this data.  
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
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  
## 4  2020     1    20 인천      1  
## 5  2020     1    20 광주      0  
## 6  2020     1    20 대전      0
```

(1)

```
df %>% group_by(year) %>% summarise(sum(cases))
```

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

(2)

```
df %>%  
filter(year==2020 & month==2 & 1 <= day & day <= 15) %>%  
group_by(prov) %>%  
summarise(sum(cases))
```

```
## # A tibble: 18 x 2  
##   prov   `sum(cases)`  
##   <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  
## 18 충북            0
```

따라서 2020년 2월 1일부터 202년 2월 15일까지 가장 많은 확진자가 발견된 지역은 경기이다.

(3)

```
df %>%  
filter(year==2020 & month==2 & 16 <= day & day <= 29) %>%  
group_by(prov) %>%  
summarise(sum(cases))
```

```
## # A tibble: 18 x 2  
##   prov   `sum(cases)`  
##   <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  
## 18 충북            10
```

따라서 2020년 2월 16일부터 2020년 2월 29일까지 가장 많은 확진자가 발견된 지역은 대구이다.