

R입문 기말고사

이주영 (201919281)

1번

(1)

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

(2)

```
x1 <- c()
for (i in 1:1000) x1[i] = sin(2*pi*i/1000)

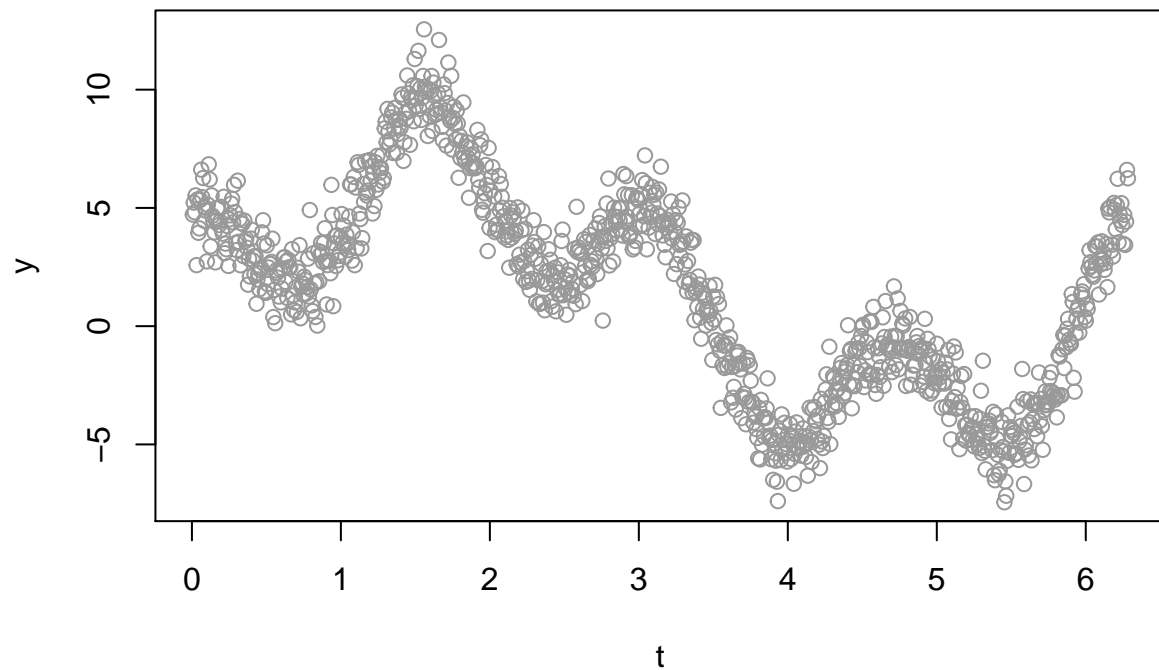
x2 <- c()
for (i in 1:1000) x2[i] = cos(8*pi*i/1000)

y = c()
for (i in 1:1000) y[i] = 1.5 + 5*x1[i] + 3*x2[i] + epsilon[i]
```

(3)

```
t<-c()
for (i in 1:1000) t[i]=2*pi*i/1000

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
```

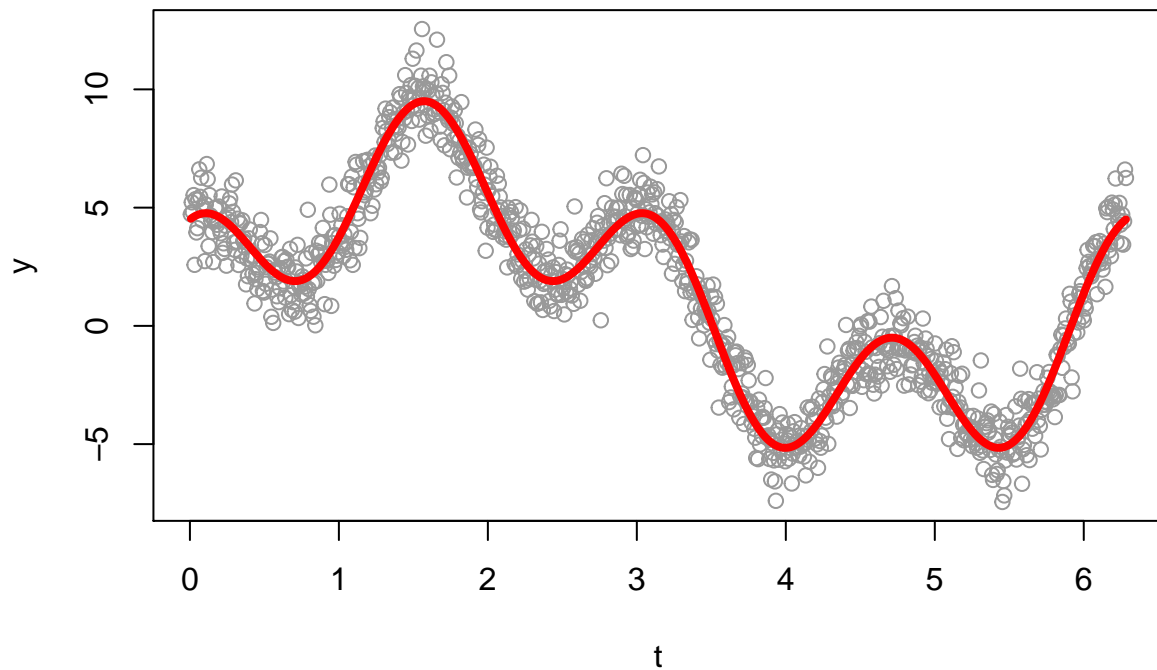
(5)

```
B=rbind(1.5,5,3)
B
```

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

```
## [3,] 3.0
x=X%%B
x_=x[,1]

plot(t,y, col='gray60')
lines(t,x_, col='red', lwd=4)
```



(6)

```
Y=cbind(y)
B_hat = solve(t(X)%%X) %% t(X) %% Y
B_hat
```

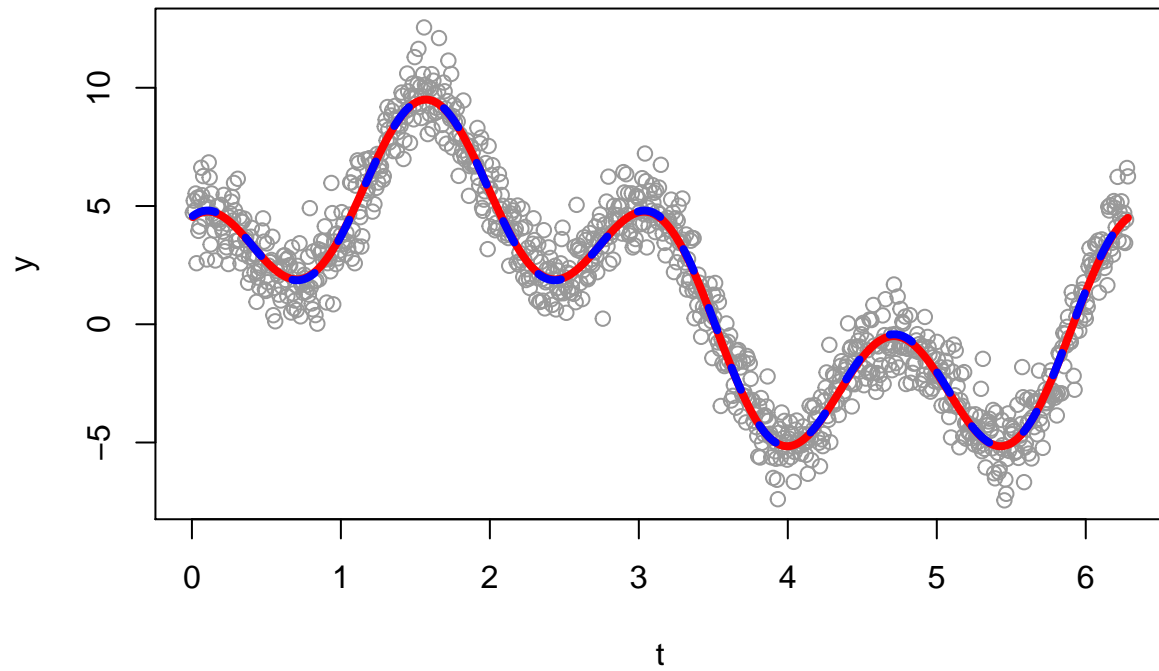
```
##          y
## 1.516166
## x1 4.973542
## x2 3.033198
```

(7)

```
x=X%%B
x_=x[,1]

hat_=X %% B_hat
XB_hat=hat_[:,1]

plot(t,y, col='gray60')
lines(t,x_, col='red', lwd=4)
lines(t,XB_hat, lty=2, col='blue', lwd=4)
```

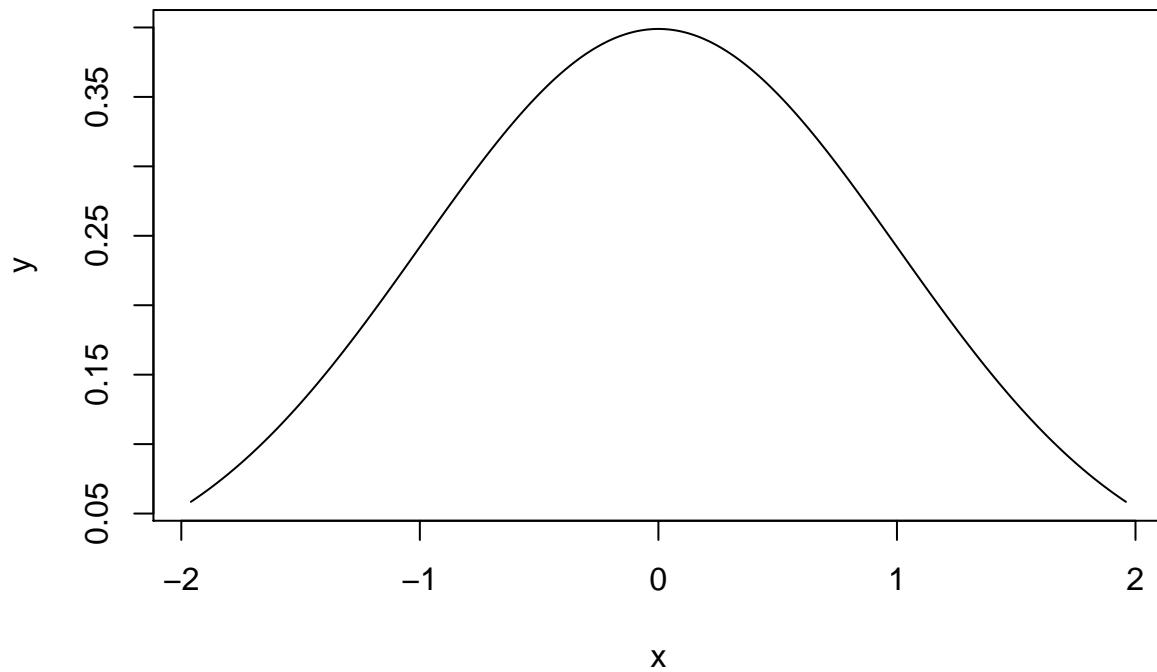


2번

(1)

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

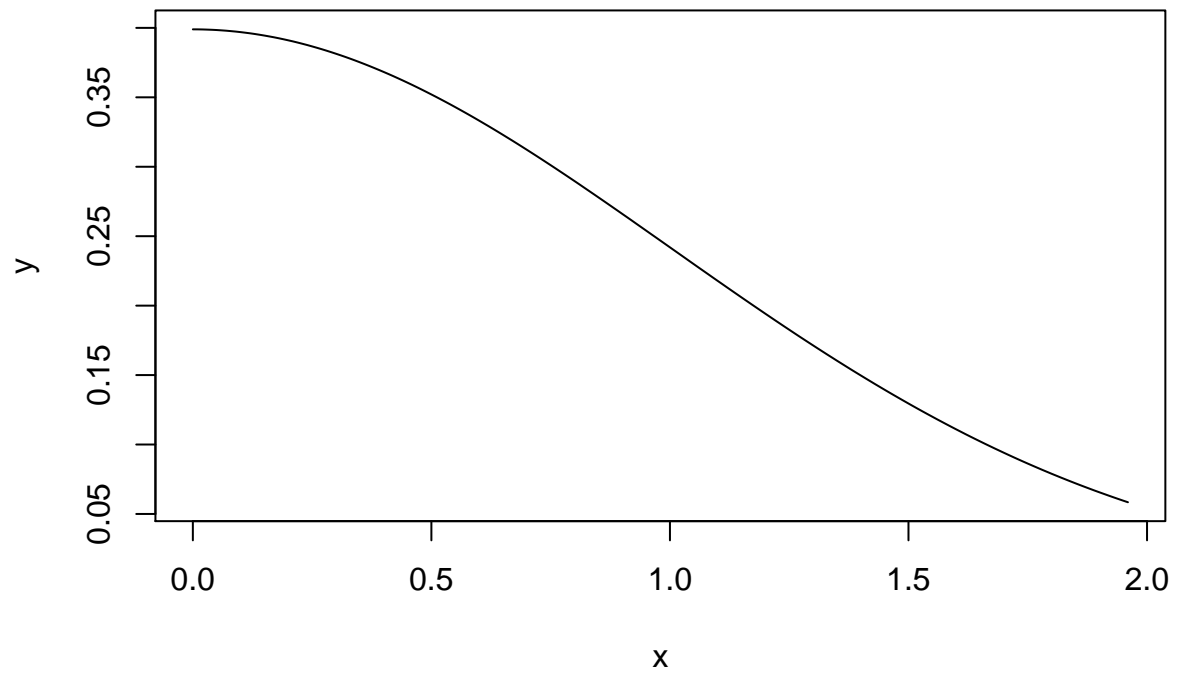
plot(x,y,type='l')
```



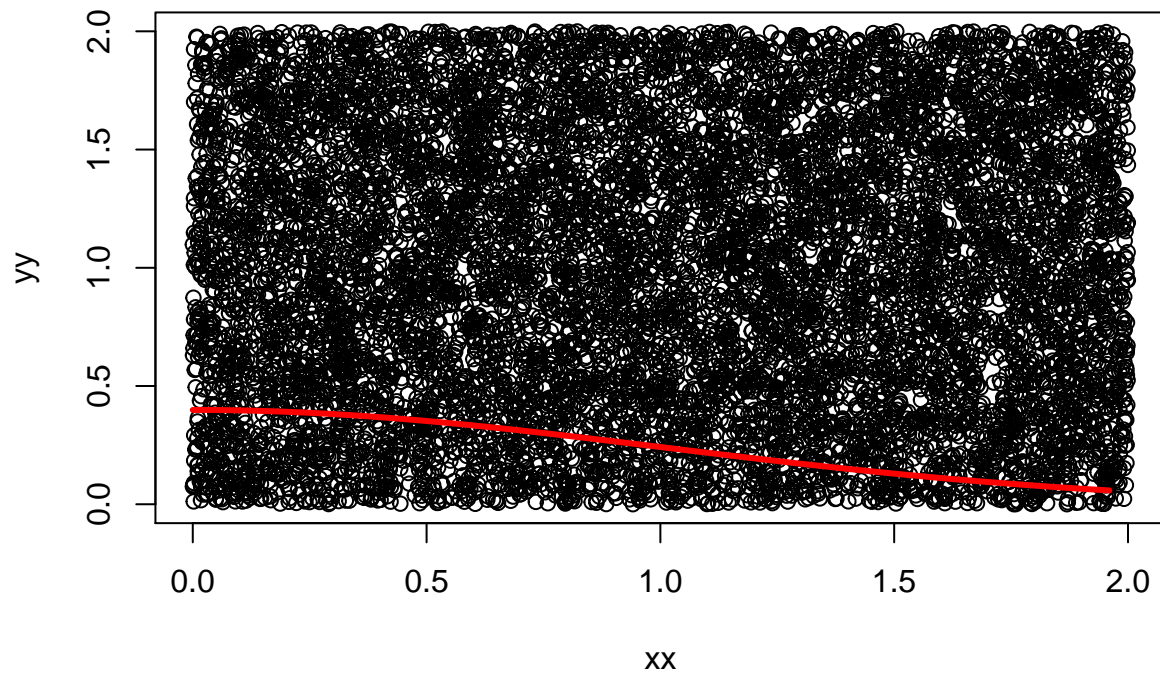
- 그래프가 대칭이므로 0부터 1.96의 넓이를 구한 뒤 2배를 해주자.

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

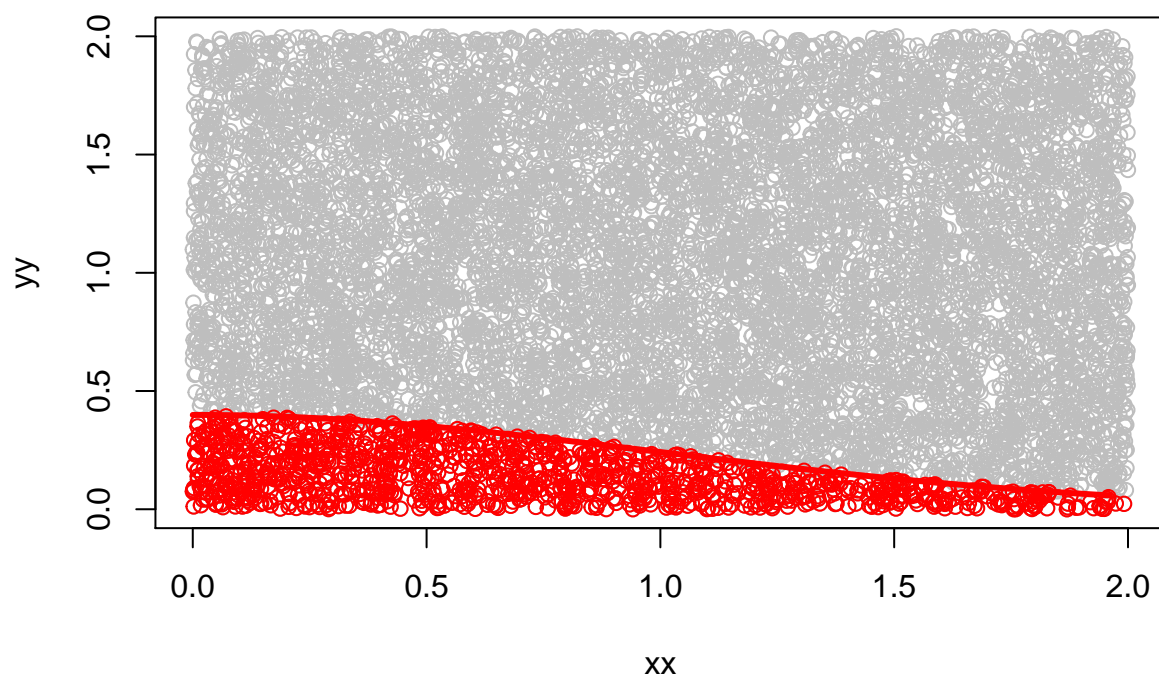
plot(x,y,type='l')
```



```
xx=runif(10000)*2  
yy=runif(10000)*2  
  
plot(xx,yy)  
lines(x,y,col='red', lwd=3)
```



```
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')
```



```
sum(tst)*2
```

```
## [1] 2360
```

```
sum(tst)*2/10000 *4
```

```
## [1] 0.944
```

(2)

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
```

```
## v tibble  3.1.2      v dplyr   1.0.6
```

```
## v tidyr   1.1.3      v stringr 1.4.0
```

```
## v readr   1.4.0      v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```



```
## Warning: package 'tidyr' was built under R version 4.0.5
## Warning: package 'readr' was built under R version 4.0.5
## Warning: package 'purrr' was built under R version 4.0.5
## Warning: package 'dplyr' was built under R version 4.0.5
## Warning: package 'stringr' was built under R version 4.0.5
## Warning: package 'forcats' was built under R version 4.0.5
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
v1=rnorm(1000)
t2=tibble(v1) %>% filter(v1>-1.96, v1<1.96)
nrow(t2)

## [1] 948
```

3번

Type A

```
rslt<-c()
for (i in 1:17) rslt[i]=sum(rbinom(1000, size=20, 0.5)==i) * sum(rbinom(1000, size=20, 0.5)==i)
sum(rslt)
```

```
## [1] 12
```

```
sum(rbinom(1000, size=20, 0.5)==18) + sum(rbinom(1000, size=20, 0.5)==19) + sum(rbinom(1000, size=20, 0.5)==20)
```

```
## [1] 603
```

Type B

```
sum(rbinom(1000, size=20, 0.5)==13) + sum(rbinom(1000, size=20, 0.5)==14) + sum(rbinom(1000, size=20, 0.5)==15)
```

```
## [1] 142
```

4번

```
library(tidyverse)

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

##
## -- Column specification -----
## cols(
##   year = col_double(),
##   month = col_double(),
##   day = col_double(),
##   prov = col_character(),
##   cases = col_double()
## )
```

```
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)

```
df2 = df %>% filter(year==2020)
df2['cases'] %>% sum()
```

```
## [1] 60726
```

```
df3 = df %>% filter(year==2021)
df3['cases'] %>% sum()
```

```
## [1] 396886
```

(2)

```
df3= df %>% filter(year==2020, month==2, day>=1, day<=15)

df3 %>% group_by(prov) %>% summarise(sum_=sum(cases))
```

```
## # 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
## 18 충북        0
```

- 경기도가 가장 많은 확진자가 발생했다.

(3)

```
df4= df %>% filter(year==2020, month==2, day>=16, day<=29)

df4 %>% group_by(prov) %>% summarise(sum_=sum(cases))
```

```
## # 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
## 18	충북	10

- 대구가 가장 많은 확진자가 발생했다.