R입문 기말고사

김현아

12/21/2021

1번

(1)

```r
epsilon= rnorm(100)
head(epsilon)
```

```
## [1] -0.6443916 -1.6630079 -2.3675968  1.2785122  1.0095531  0.6701259
```

(2)

```r
t<- 2*(1:1000)*pi/1000
head(t)
```

```
## [1] 0.006283185 0.012566371 0.018849556 0.025132741 0.031415927 0.037699112
```

```r
x1<- sin(t)
head(x1)
```

```
## [1] 0.006283144 0.012566040 0.018848440 0.025130095 0.031410759 0.037690183
```
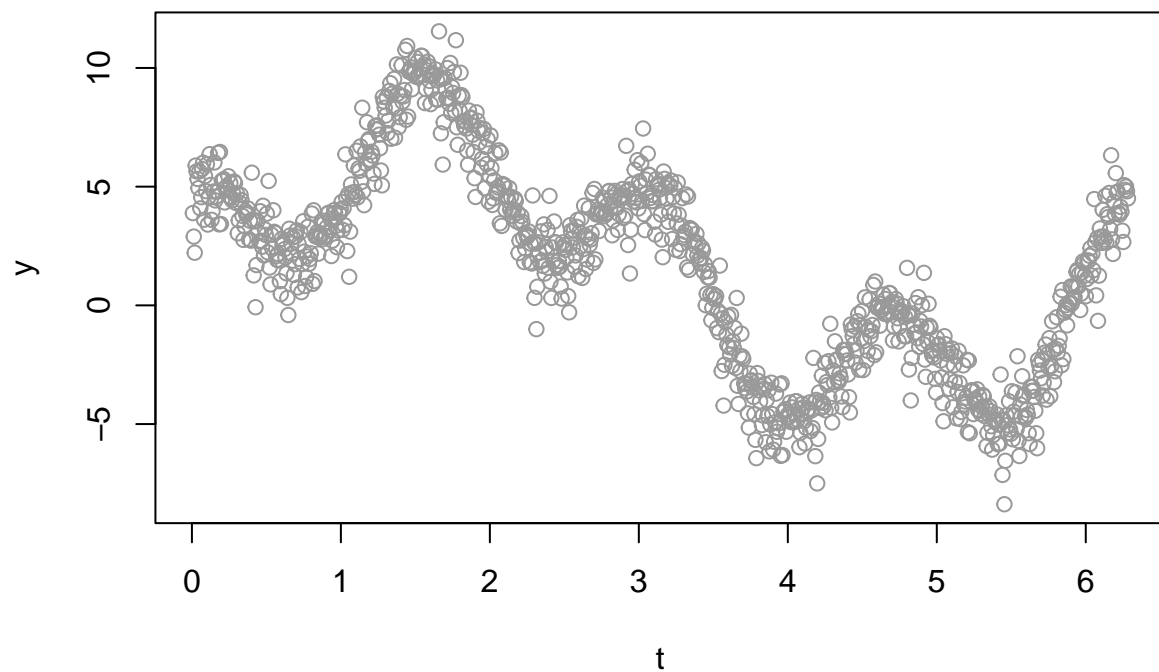
```r
x2<- cos(4*t)
head(x2)
```

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

(3)

```r
y= 1.5+ 5*x1+ 3*x2+ epsilon
head(y)
```

```
## [1] 3.886077 2.896033 2.218122 5.889016 5.642951 5.324532
```

```r
plot(t,y,col='gray60')
```

(4)

```
tmp= rep(1,1000)
X= cbind(tmp,x1,x2)
print(head(X))

##      tmp          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
```
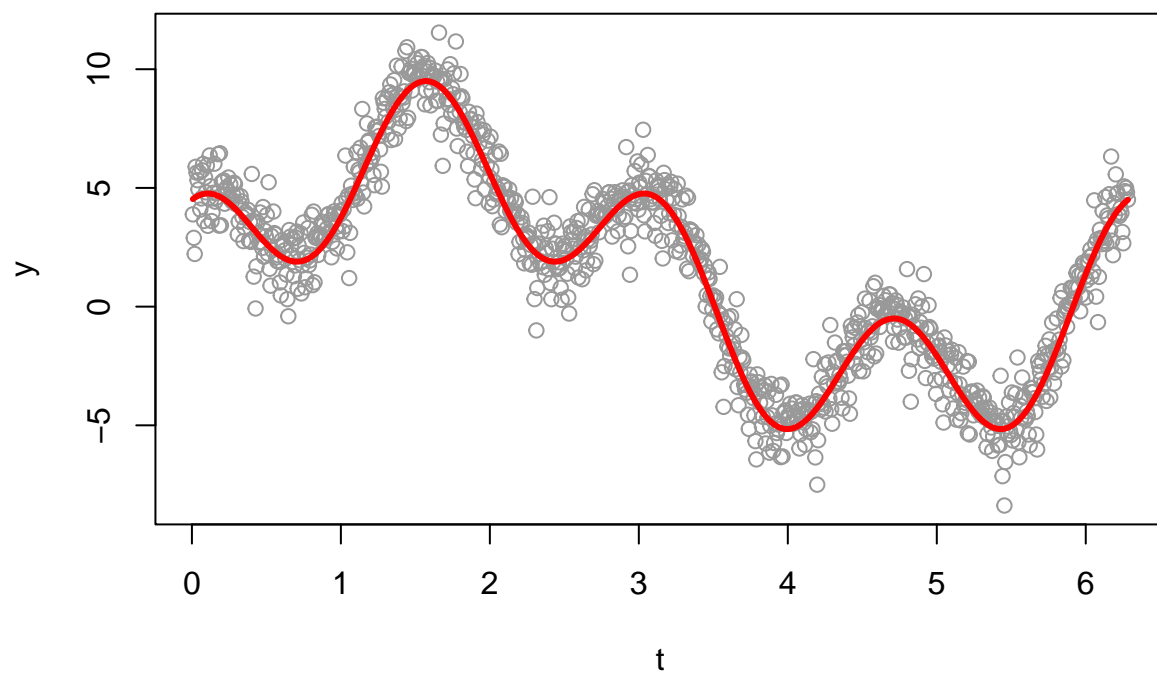
(5)

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

##      [,1]
## [1,]  1.5
```

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

```
Xbeta= X %*% beta
head(Xbeta)
```

```
##             [,1]
## [1,] 4.530468
## [2,] 4.559041
## [3,] 4.585719
## [4,] 4.610504
## [5,] 4.633398
## [6,] 4.654406
```

```
plot(t,y,col='gray 60')
lines(t,Xbeta, col='red', lwd=3)
```
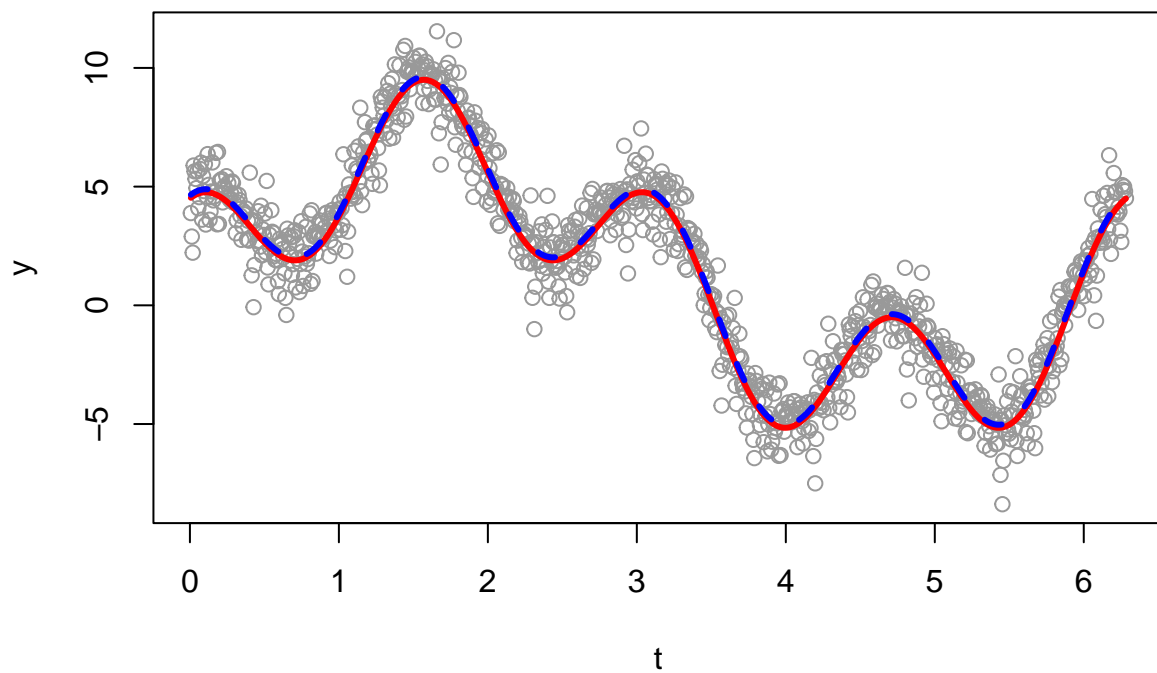

(6)

```
beta_hat= solve(t(X) %*% X) %*% t(X) %*% y
beta_hat
```

```
##             [,1]
```

```
## tmp 1.628127
## x1  5.000000
## x2  3.000000
```
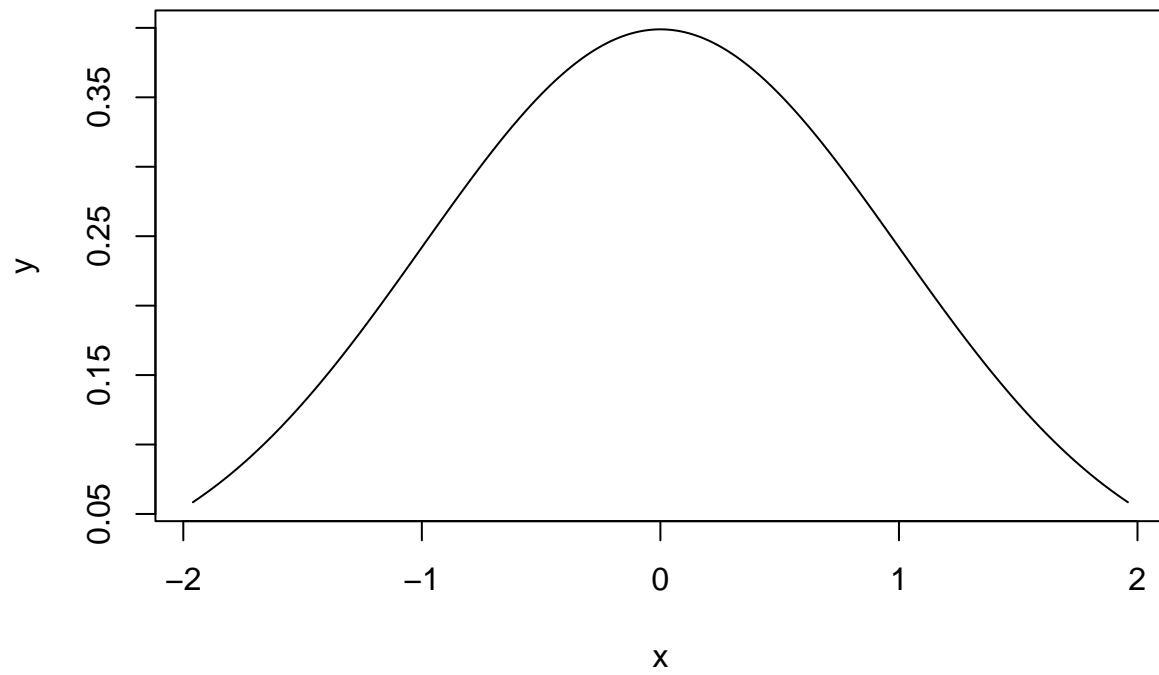
(7)

```
Xbeta_hat= X %*% beta_hat
plot(t,y,col='gray 60')
lines(t,Xbeta, col='red', lwd=3)
lines(t,Xbeta_hat,col='blue', lty=2, lwd=3)
```
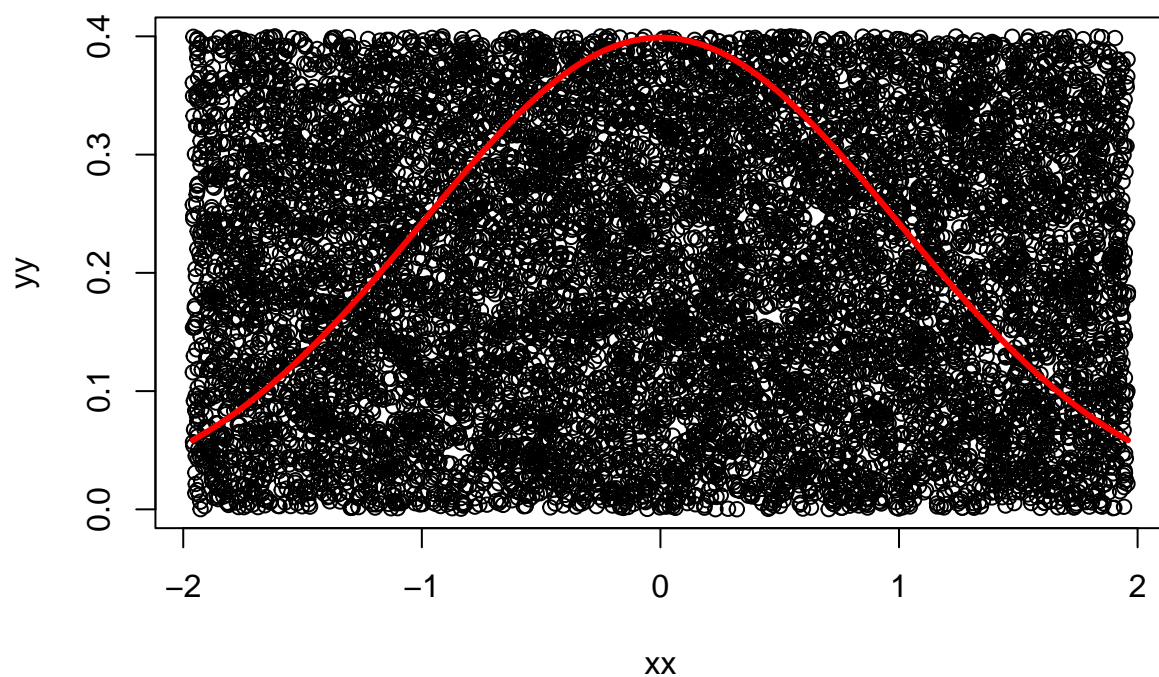

2번

(1)

```
x=seq(-1.96,1.96,0.01)
y=(1/sqrt(2*pi))*exp(-0.5*x^2)
plot(x,y,type='l')
```
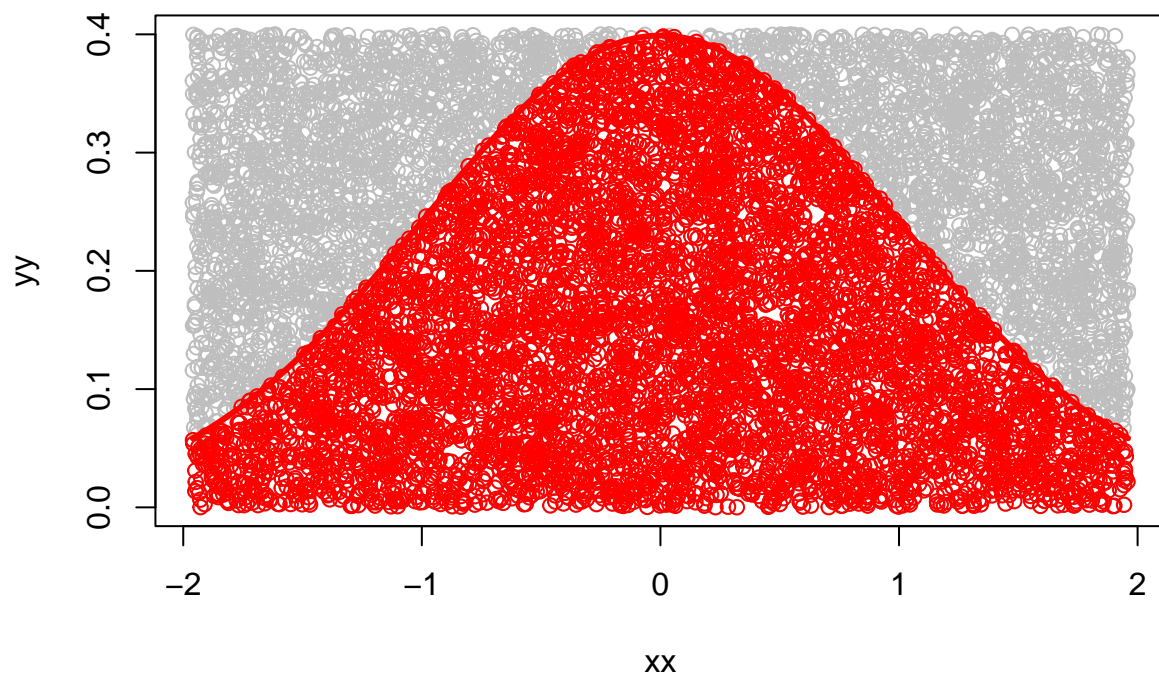
```
xx=runif(10000,-1.96,1.96)
yy=runif(10000,0,0.40)
plot(xx,yy)
plot(xx,yy)
lines(x,y,type='l',col='red',lwd=3)
```

```r
test= function(xx,yy){
  yy< (1/sqrt(2*pi))*exp(-0.5*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=3)
points(xx[tst],yy[tst],col='red')
```

```
sum(tst)
```

```
## [1] 6152
```

```
sum(tst)/10000 * 1.568
```

```
## [1] 0.9646336
```

(2)

```
arr_rnorm= rnorm(1000)
count=0
for(i in 1:1000){
  if(arr_rnorm[i]< 1.96 && arr_rnorm[i] > -1.96)
    count=count+1
}
count
```

```
## [1] 945
```

3번

(B)

```r
prob<- 0
for(i in 20:13){
  prob<- prob + dbinom(i,20,0.5)
  print(prob)
}
```

```
## [1] 9.536743e-07
## [1] 2.002716e-05
## [1] 0.0002012253
## [1] 0.001288414
## [1] 0.005908966
## [1] 0.02069473
## [1] 0.05765915
## [1] 0.131588
```

따라서 8번참가자가 생존할 확률 0.131588

4번

```r
library(tidyverse)
```

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

## -- Attaching packages ----------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.1      v forcats 0.5.1
## Warning: 패키지 'ggplot2'는 R 버전 4.1.2에서 작성되었습니다

## Warning: 패키지 'tibble'는 R 버전 4.1.2에서 작성되었습니다

## Warning: 패키지 'tidyr'는 R 버전 4.1.2에서 작성되었습니다

## Warning: 패키지 'readr'는 R 버전 4.1.2에서 작성되었습니다

## Warning: 패키지 'purrr'는 R 버전 4.1.2에서 작성되었습니다

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

## Warning: 패키지 'stringr'는 R 버전 4.1.2에서 작성되었습니다

## Warning: 패키지 'forcats'는 R 버전 4.1.2에서 작성되었습니다

## -- Conflicts -------------------------------------- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
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.
```

(1)

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

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

(2)

```
df %>% filter(year==2020, month==2, day<16) %>%
  group_by(prov) %>% summarise(cases_sum= sum(cases)) %>%
  arrange(desc(cases_sum))
```

```
## # A tibble: 18 x 2
##     prov  cases_sum
##    <chr>      <dbl>
##  1 경기          9
##  2 서울          5
##  3 광주          2
##  4 전남          1
##  5 강원          0
##  6 검역          0
##  7 경남          0
##  8 경북          0
##  9 대구          0
## 10 대전          0
## 11 부산          0
```

```
## 12 세종           0
## 13 울산           0
## 14 인천           0
## 15 전북           0
## 16 제주           0
## 17 충남           0
## 18 충북           0
```

(3)

```
df %>% filter(year==2020, month==2, 15<day) %>%
  group_by(prov) %>% summarise(cases_sum= sum(cases)) %>%
  arrange(desc(cases_sum))
```

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