

R입문 기말고사

김채연(202115791)

12/21/2021

1번 문제

(1)

```
A <- rnorm(1000)
```

```
print(head(A))
```

```
## [1]  1.87018875 -0.06286546 -0.09619722  0.13033713  0.68665119 -0.03443760
```

(2)

```
t <- c()
```

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

```
x1 <- c()
```

```
for(i in 1:1000) x1[i] <- sin(2*t[i])
```

```
print(head(x1))
```

```
## [1] 0.01256604 0.02513010 0.03769018 0.05024432 0.06279052 0.07532681
```

```
x2 <- c()
```

```
for(i in 1:1000) x2[i] <- cos(4*t[i])
```

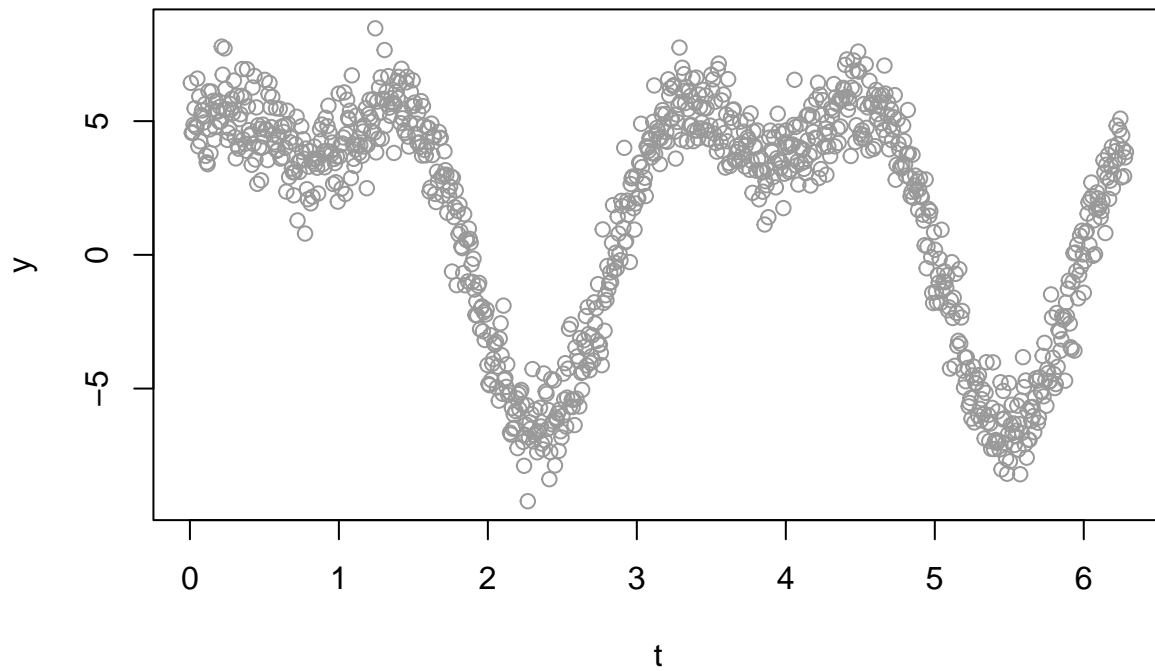
```
print(head(x2))
```

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

(3)

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

plot(t,y,col="gray60")
```



(4)

```
x=cbind(1,x1,x2)
print(head(x))
```

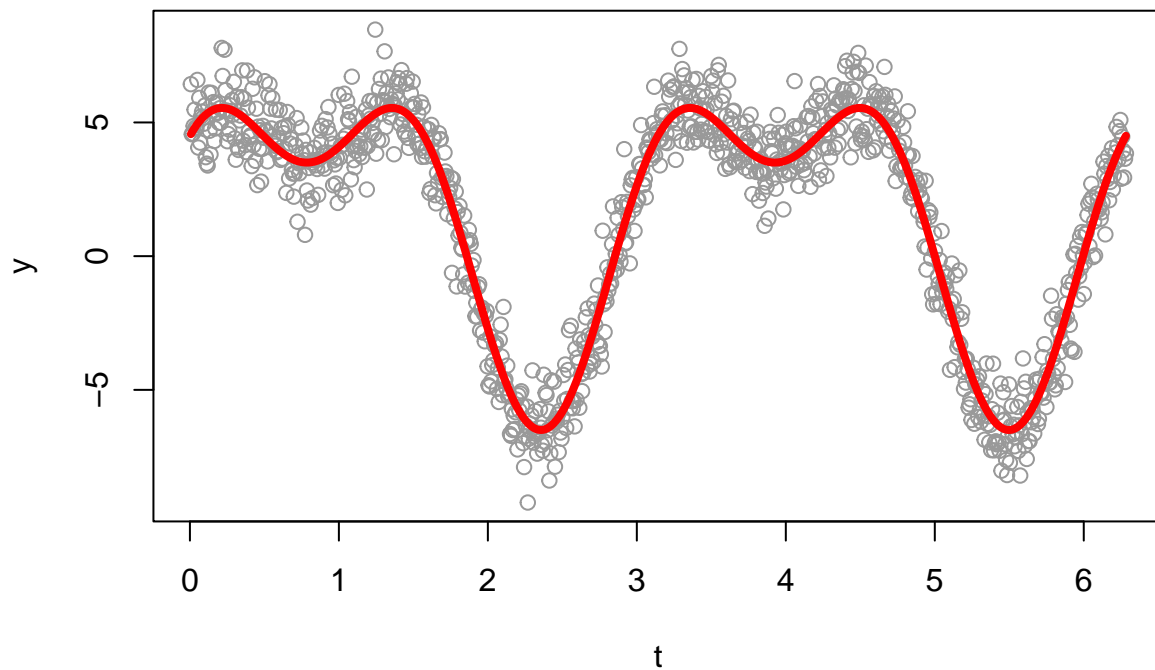
```
##           x1           x2
## [1,]  1 0.01256604 0.9996842
## [2,]  1 0.02513010 0.9987370
## [3,]  1 0.03769018 0.9971589
## [4,]  1 0.05024432 0.9949510
## [5,]  1 0.06279052 0.9921147
## [6,]  1 0.07532681 0.9886517
```

(5)

```
b <- c(1.5,5,3)
xb <- x%*%b
print(head(xb))
```

```
##           [,1]
## [1,] 4.561883
## [2,] 4.621861
## [3,] 4.679928
## [4,] 4.736075
## [5,] 4.790297
## [6,] 4.842589
```

```
plot(t,y,col="gray60")
lines(t, xb, col="red",lwd=4)
```



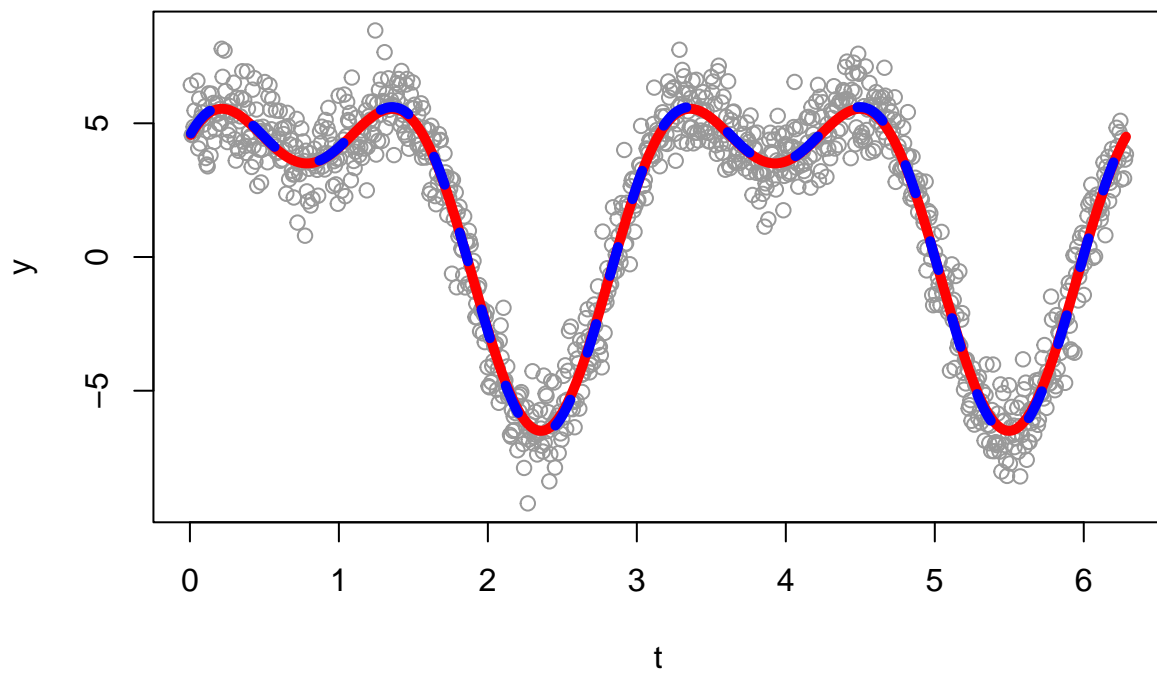
(6)

```
b_ <- solve(t(x)%*%x)%*%t(x)%*%y  
print(b_)
```

```
##      [,1]  
## 1.498348  
## x1 5.066977  
## x2 3.058352
```

(7)

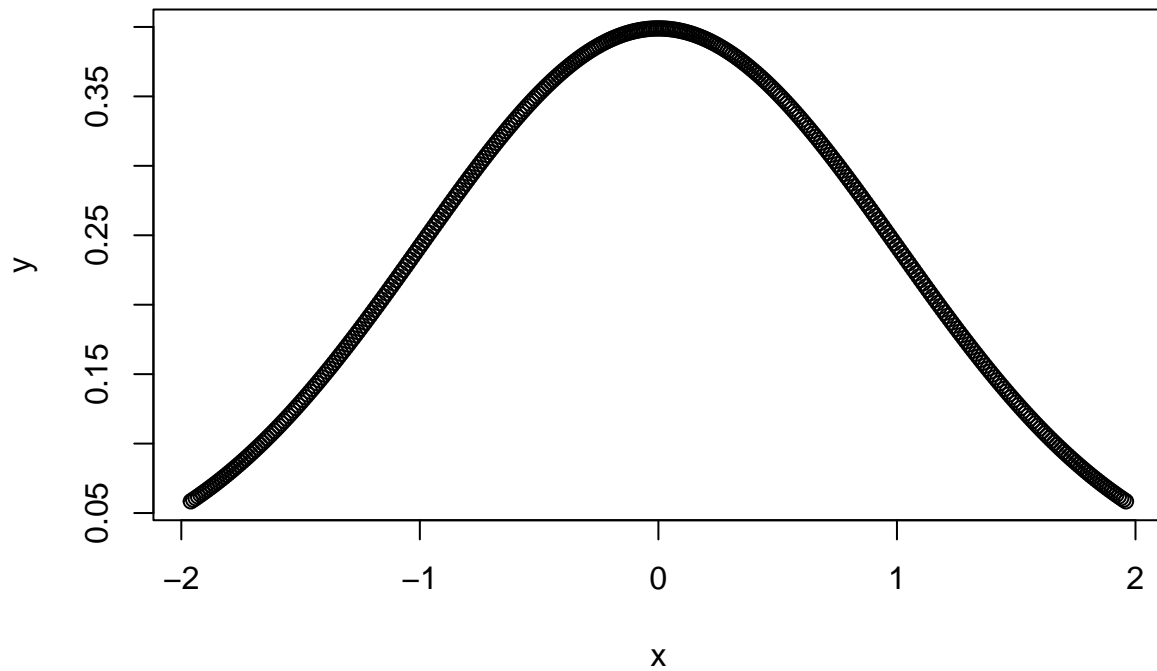
```
xb_ <- x%*%b_  
  
plot(t,y,col="gray60")  
lines(t, xb, col="red",lwd=5)  
lines(t,xb_,col="blue",lty=2,lwd=5)
```



2번문제

(1)

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



```
xx= runif(10000,-1.96,1.96)
yy= runif(10000,0.05,0.40)
```

```
tst=sum(yy < exp(-1/2*xx^2)/sqrt(2*pi))
```

$$\frac{s}{(1.96 + 1.96) * (0.40 - 0.05)} \approx \frac{tst}{10000}$$

```
tst/10000 * (1.96*2*(0.4-0.05))
```

```
## [1] 0.7474656
```

(2)

```
a <- rnorm(1000)
```

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

```
## [1] 949
```

3번 문제

type A

-10번과 9번이 건널 다리의 개수

```
rslt_ <- c()
for (i in 1:1000){
  x=sum(cumprod(rbinom(20,size=1,0.5)))
  xx_=19-x
  if (xx_>0) y= sum(cumprod(rbinom(xx_,size=1,0.95))) else y=0
  rslt_[i]=x+y
}
```

-남은 다리를 8번이 모두 건널 확률

```
rslt<- c()
for( i in 1:1000){
  if(19-rslt_[i]==0) rslt[i]=1 else rslt[i]=0.5^(19- rslt_[i])
}
sum(rslt)/1000
```

```
## [1] 0.4116848
```

type B

```
B <- c()
for(i in 1:10){
  B[i]= sum(rbinom(1000,size=20,p=0.5)== i)
}
```

```
sum(B[1:7])/1000
```

```
## [1] 0.134
```

따라서 8번 참가자는 Type A의 경우에 살아남을 확률이 더 높다.

4번문제

```
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()
```


(1)

```
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
```

-2020년의 확진자 총합

```
df %>% filter(year==2020) -> df2020  
sum(df2020$cases)
```

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

-2021년의 확진자 총합

```
df %>% filter(year==2021) -> df2021  
sum(df2021$cases)
```

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

(2)

```
df %>%  
group_by(prov)%>%  
filter(year==2020 & month==2 & 1<=day & day<= 15 )%>%  
summarise(cases_sum=sum(cases,na.rm=TRUE))
```

```
## # A tibble: 18 x 2  
##   prov cases_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
```

따라서 가장 많은 확진자가 발견된 지역은 경기이다.

(3)

```
df %>%  
group_by(prov)%>%  
filter(year==2020 & month==2 & 16<=day & day<= 29 )%>%  
summarise(cases_sum=sum(cases,na.rm=TRUE))
```

```
## # A tibble: 18 x 2  
##   prov cases_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
```

따라서 가장 많은 확진자가 발견된 지역은 대구이다.