

R 입문 기말고사

이태관(202121238)

12/21/2021

1번

```
##(1)
e = rnorm(1000)
##(2)
t = c()
x_1 = c()
x_2 = c()
y = c()
for(i in 1:1000) t[i] = ((2*pi*i)/1000)
for(i in 1:1000) x_1[i] = sin(t[i])
for(i in 1:1000) x_2[i] = cos(4*t[i])

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

##(4)
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

##(5)
B = rbind(1.5,5,3)
V = as.vector(X %*% B)
```

```

head(V)

## [1] 4.530468 4.559041 4.585719 4.610504 4.633398 4.654406
lines(t,V,col='red', lwd=5)

##(6)
b_h = solve(t(X) %*% X) %*% t(X) %*% y
print(B)

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

print(b_h)

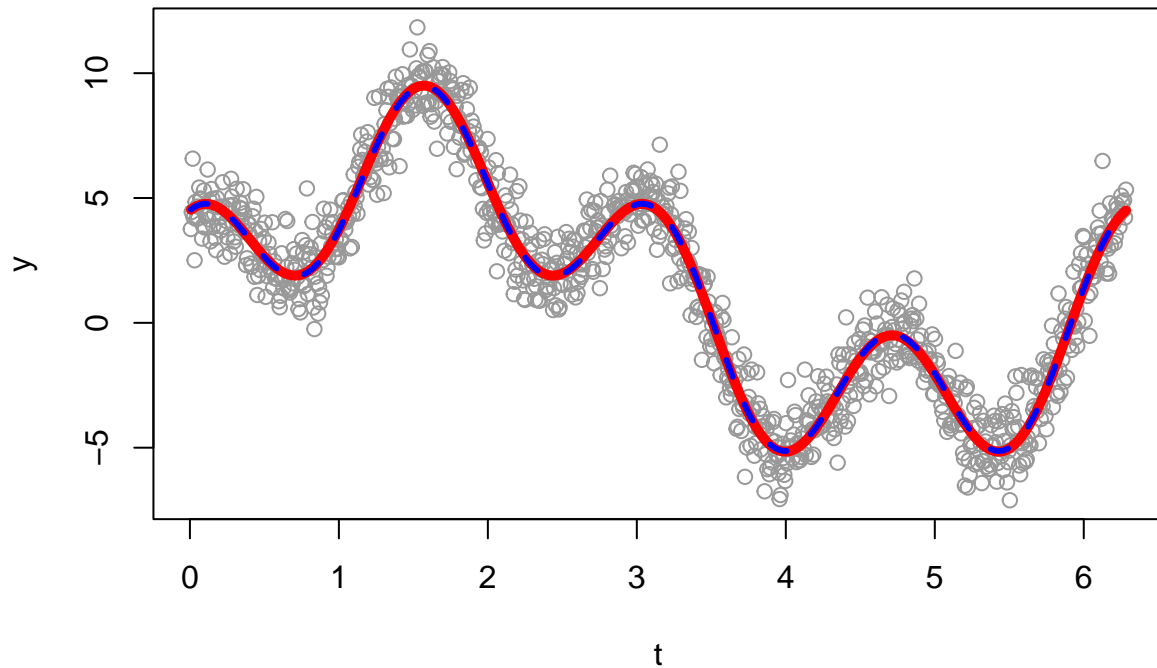
##      [,1]
##      1.504258
## x_1 4.963066
## x_2 3.005636

## 둘이 거의 동일함을 알 수 있다.

##(7)
V_2 = as.vector(X %*% b_h)
head(V_2)

## [1] 4.540129 4.568464 4.594901 4.619441 4.642087 4.662844
lines(t,V_2,col='blue', lwd=3, lty="dashed")

```



2번

```
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)
x=seq(from=-1.96,to=1.96, by=0.01)
y_1= (1/sqrt(2*pi))*exp((-1/2)*(x^2))
x_one = runif(5000)*-2
x_two = runif(5000)*2
xx = c(x_one,x_two)
yy = runif(10000)*(0.4)
plot(xx,yy)
lines(x,y_1,col='red', lwd=3)
test = function(xx,yy){
  yy < abs(1/sqrt(2*pi))*exp((-1/2)*(xx^2))
}
print(c(xx[1],yy[1]))

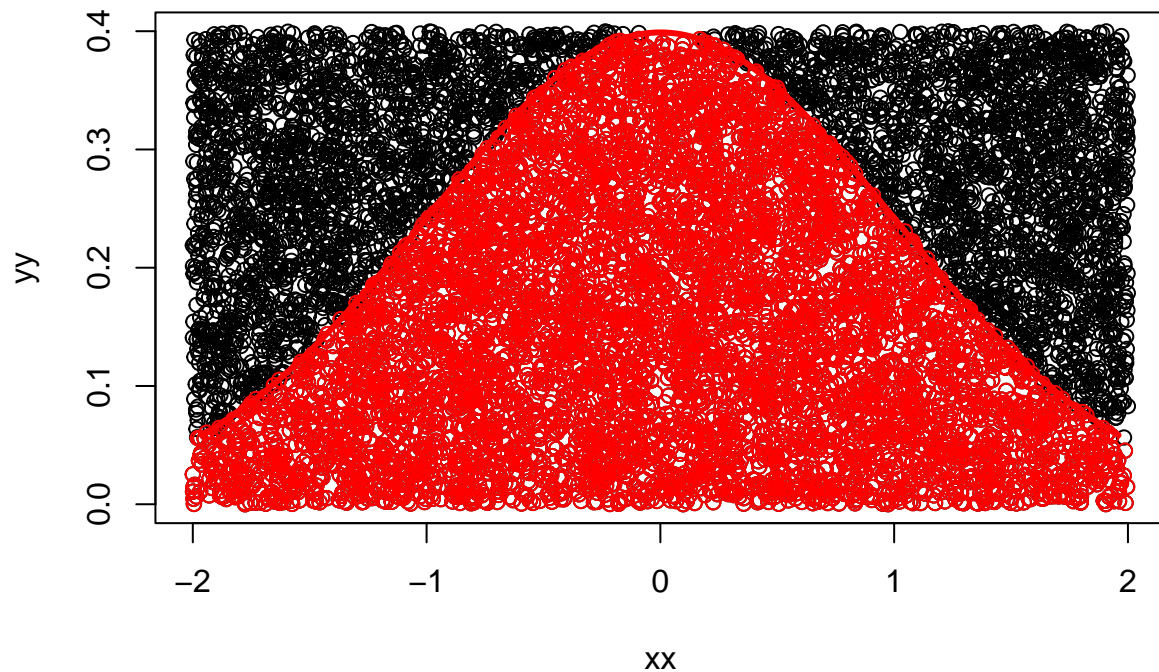
## [1] -0.61168174  0.03964701

test(xx[1],yy[1])

## [1] TRUE

tst = c()
##제대로 작동한다.

for (i in 1:10000) tst[i] = test(xx[i],yy[i])
points(xx[tst],yy[tst],col='red')
```



```
sum(tst)
```

```
## [1] 6031
```

```
sum(tst)/10000 * 1.6
```

```
## [1] 0.96496
```

약 0.95임을 알 수 있다.

##(2)

```
rndm = rnorm(1000)
rndm = as_tibble(rndm)
colnames(rndm) = c('X1')
rndm %>% filter(X1>-1.96 & X1<1.96) %>% count
```

```
## # A tibble: 1 x 1
```

```
##       n
```

```
##   <int>
```

```
## 1    964
```

##약 940~950개가 있음을 알 수 있다.

3번

```
##Type A
library(tidyverse)
ARR = c('N1','N2','N3','N4','N5','N6','N7','N8','PRO','N9')
SURV = 10
PLAYER = ARR[SURV]
STAGE = 0
PROB = 0.5
TOSSRSLT = NA
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(){
  if(SURV == 0){
    return()
  }
  PROB <- 0.5+ (PLAYER=='PRO')*0.45
  TOSSRSLT <- toss(PROB)
  if(TOSSRSLT==FALSE) SURV <- SURV - 1
  STAGE <- STAGE + 1
  PLAYER <- ARR[SURV]
}

gogo = function() for(i in 1:20) go()

simulate_once = function(){
  reset()
  gogo()
  return(record())$SURV
}

simrslt = c()
for(i in 1:10000) simrslt[i] = simulate_once()
mean(simrslt)

## [1] 5.6056
```

```
sum(simrslt>=8) ## 생존자가 8명 이상이어야 8번이 산다.
```

```
## [1] 4472
```

```
## Type B
```

```
library(tidyverse)
ARR = c('N9','PRO','N8','N7','N6','N5','N4','N3','N2','N1')
SURV = 10
PLAYER = ARR[SURV]
STAGE = 0
PROB = 0.5
TOSSRSLT = NA
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(){
  if(SURV == 0){
    return() ##생존자가 0명이 될 경우 go를 거치지 않고 반환
  }
  PROB <- 0.5+ (PLAYER=='PRO')*0.45
  TOSSRSLT <- toss(PROB)
  if(TOSSRSLT==FALSE) SURV <- SURV - 1
  STAGE <- STAGE + 1
  PLAYER <- ARR[SURV]
}

gogo = function() for(i in 1:20) go()

simulate_once = function(){
  reset()
  gogo()
  return(record())$SURV
}

simrslt = c()
for(i in 1:10000) simrslt[i] = simulate_once()
mean(simrslt)
```

```
## [1] 1.9026
```

```
sum(simrslt>=3) ##생존자가 3명 이상이면 8번이 살아남은 것이다.
```

```
## [1] 1357
```

```
## Type A 의 경우가 8번 생존자가 더 많이 살아남을 수 있다.
```

4번

```
library(tidyverse)
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(confirm = sum(cases))
```

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

```
##(2)
df %>%
```



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

```
## # A tibble: 18 x 2
##   prov confirm
##   <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
```

##경기에서 가장 많이 발생했음을 알 수 있다.

##(3)

```
df %>%
filter(year == 2020 & month == 2 & day > 15 & day < 30 )%>%
group_by(prov) %>%
summarise(confirm = sum(cases))
```

```
## # A tibble: 18 x 2
##   prov confirm
##   <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
```

##대구에서 가장 많이 발생했음을 알 수 있다.

###5번

##(1)

```
cumprod(c(7,6,5,4,3,2,1))
```

```
## [1]      7    42   210   840  2520  5040  5040
```

##(3)

```
library(tidyverse)
random = rnorm(1000)
random = as_tibble(random)
colnames(random) = c('P1')
random %>% filter(P1>0) %>% count
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   497
```

##(4)

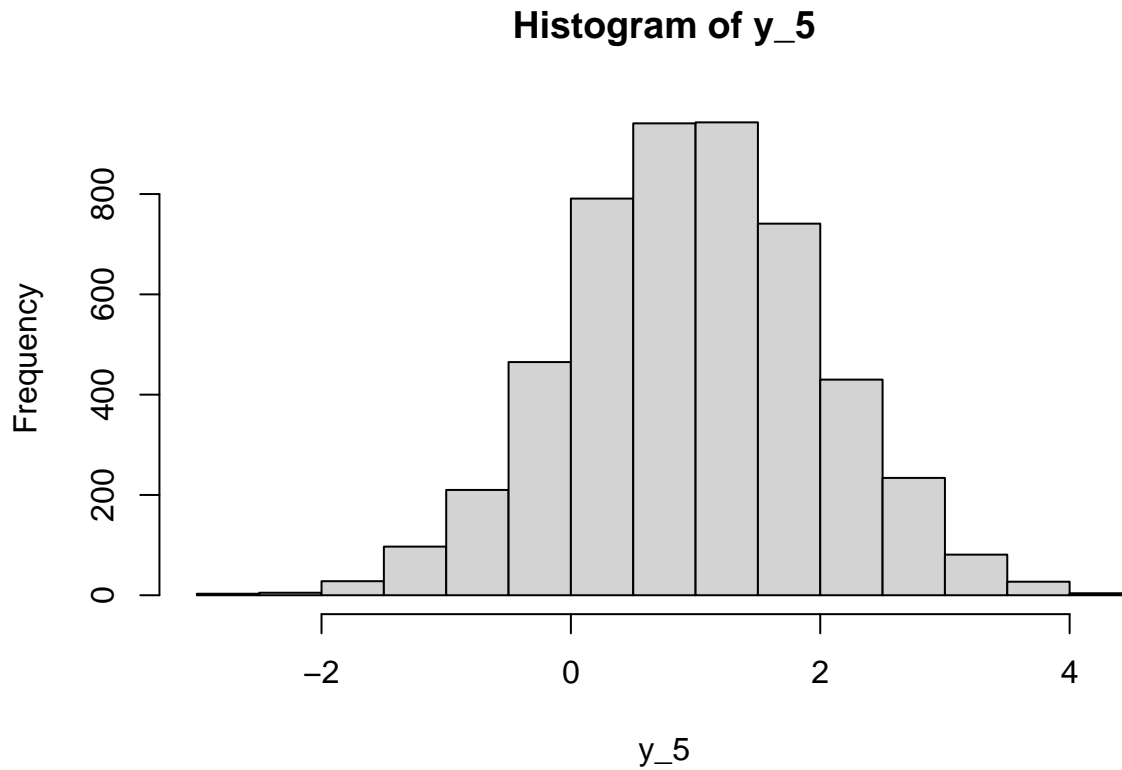
```
rbinom(1000,size=1, prob=0.2) %>% mean
```

```
## [1] 0.209
```

##약 200개가 포함된다는 것을 알 수 있다.

##(5)

```
y_5 = rnorm(5000,1,1)
hist(y_5)
```



```
##(6)
x_6 = -1000:1000/1000
sum_4=0
for(i in 1:1000){
  sum_4= sum_4 + x_6[2*i]
}
print(sum_4)
```

```
## [1] -9.325873e-15
```

```
##(7)

##(8)
calculate = function(x){
  if(x>0){
    return(log(x))
  }else{
    if(x==0){
      print("음의 무한대")
    }else{
      print("x는 양수이어야 합니다.")
    }
  }
}
```

```
}  
}
```

```
##(9)
```

```
df=tibble(X1=rnorm(100),X2=rnorm(100))  
head(df)
```

```
## # A tibble: 6 x 2  
##       X1      X2  
##   <dbl> <dbl>  
## 1 -2.10  -0.684  
## 2 -1.57   0.928  
## 3  1.25   1.40  
## 4  0.627  0.611  
## 5  0.0243 -0.750  
## 6 -1.04   0.330
```

```
df %>% mutate(Z=(X1^2+X2^2))
```

```
## # A tibble: 100 x 3  
##       X1      X2      Z  
##   <dbl> <dbl> <dbl>  
## 1 -2.10  -0.684  4.87  
## 2 -1.57   0.928  3.31  
## 3  1.25   1.40   3.54  
## 4  0.627  0.611  0.767  
## 5  0.0243 -0.750  0.563  
## 6 -1.04   0.330  1.18  
## 7 -0.493  -0.914  1.08  
## 8 -0.0377  0.544  0.297  
## 9 -1.46  -0.131  2.15  
## 10  2.16  -0.00971  4.68  
## # ... with 90 more rows
```

```
##6번
```

```
##(4)
```

```
MAT = rbind(c(1,2,3,4),c(2,2,1,0), c(0,1,-1,0), c(3,1,-1,3))  
MAT_1 = c(1,9,4,7)  
dim(MAT_1)=c(4,1)  
solve(MAT) %*% MAT_1
```

```
##      [,1]  
## [1,]    2  
## [2,]    3  
## [3,]   -1  
## [4,]   -1
```

##답은 $w=2$ $x=3$, $y=-1$, $z=-1$ 가 된다.