

R프로그래밍 및 실습 (Quiz-2) - 문제 및 답안

2018-05-01

1. 아래의 R코드의 의미 및 결과를 쓰시오.

```
A <- 1:4 # 1부터 4까지 정수벡터를 A에 저장
B <- c(3,5,7,9) # 벡터 (3,5,7,9) 를 B에 저장
C <- rep(2, 4) # 2를 4번 반복한 수(벡터)를 C에 저장
x <- data.frame(A, B, C) # A,B,C 세개의 벡터를 열로하는 데이터프레임을 생성
dim(x) # x의 차원(행 및 열의 수)
```

```
[1] 4 3
```

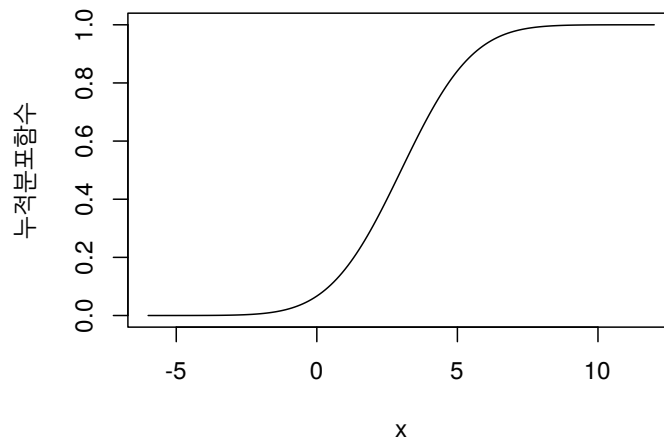
```
print(x) # x를 화면에 출력
```

```
  A B C
1 1 3 2
2 2 5 2
3 3 7 2
4 4 9 2
```

2. 아래는 모평균이 3이고 모표준편차가 2인 정규분포의 누적분포함수(cumulative distribution function)이다. 이를 그리기 위한 R 코드를 작성하라. 단, x 의 범위는 $(-6, 12)$ 이다.

```
x <- seq(-6, 12, by=0.1) # -6부터 12까지 0.1씩 증가하는 수열을 생성
y <- pnorm(x, mean=3, sd=2) # x값에 대하여 N(3, 4)의 누적분포함수값을 벡터 y로 저장
plot(x, y, type="l", ylab="누적분포함수", main="N(3, 4)의 누적분포함수")
```

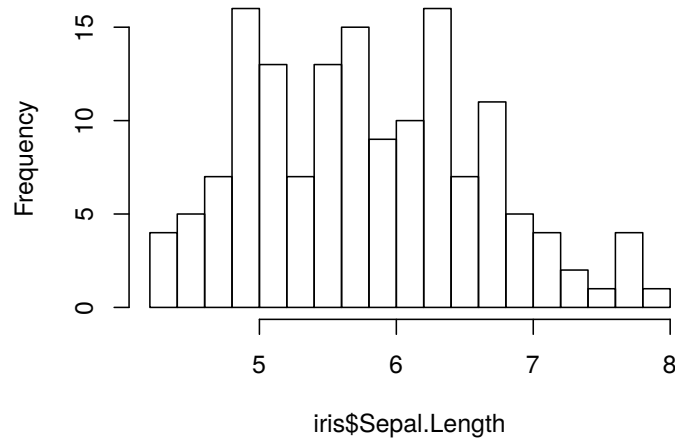
N(3, 4)의 누적분포함수



3. 아래의 코드에서 각 인수(arguments)의 의미를 해석하시오.

```
# iris$Sepal.Length: iris데이터프레임의 Sepal.Length변수값
# nclass=20: 히스토그램의 계급의 수
# main="Histogram of Sepal Length": 히스토그램의 타이틀
hist(iris$Sepal.Length, nclass=20, main="Histogram of Sepal Length")
```

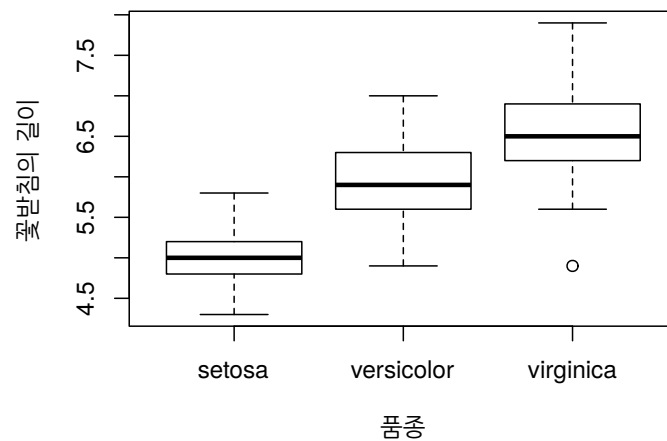
Histogram of Sepal Length



4. R에 내장된 붓꽃(iris) 데이터를 이용하여 품종(Species)별 꽃받침의 길이(Sepal.Length)에 대한 상자그림을 아래처럼 출력하려고 한다. R 코드를 작성하라.

```
boxplot(Sepal.Length~Species, data=iris, main="품종별 꽃받침의 길이에 대한 상자그림",
        xlab="품종", ylab="꽃받침의 길이")
```

품종별 꽃받침의 길이에 대한 상자그림



5. 아래의 자료는 2011년 부터 2014년까지 우리나라의 시도별 인구밀도이다. 이 자료는 아래의 코드를 실행하면 얻을 수 있다. 아래 R코드의 라인 별로 코드의 의미를 쓰시오.

지역	X2011	X2012	X2013	X2014
전국	497	499	501	503
서울	16592	16497	16402	16343
부산	4498	4464	4450	4432

지역	X2011	X2012	X2013	X2014
대구	2798	2795	2790	2784
인천	2659	2677	2715	2728
광주	2999	3012	3021	3025
대전	2820	2841	2851	2866
울산	1042	1053	1064	1073
경기	1161	1176	1193	1207
강원	89	89	89	89
충북	207	208	210	210
충남	244	249	251	253
전북	223	223	223	223
전남	144	144	143	143
경북	138	139	139	139
경남	307	308	309	311
제주	299	303	308	314

```
# csv 포맷의 파일을 읽어 데이터프레임 x를 생성
x <- read.csv("http://datamining.dongguk.ac.kr/test/pop_density.csv")
head(x) # 데이터프레임 x의 앞줄을 출력
```

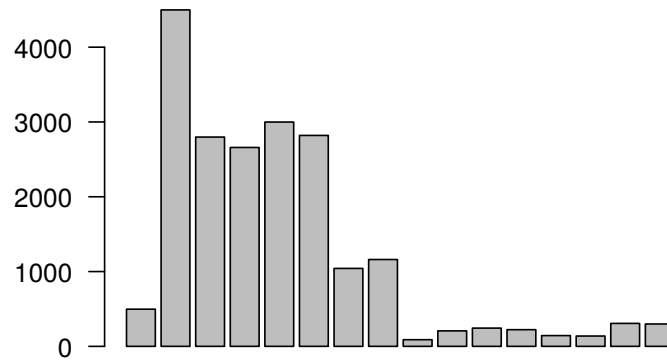
```
지역 X2011 X2012 X2013 X2014
1 전국 497 499 501 503
2 서울 16592 16497 16402 16343
3 부산 4498 4464 4450 4432
4 대구 2798 2795 2790 2784
5 인천 2659 2677 2715 2728
6 광주 2999 3012 3021 3025
```

```
names(x) # 데이터프레임 x의 변수명을 출력
```

```
[1] "지역" "X2011" "X2012" "X2013" "X2014"
```

```
# 데이터프레임 x의 변수 X2011에서 두번째 값을 제외한 나머지를 이용하여 막대그림을 생성
barplot(x$X2011[-2], main="2011년 시도별 인구밀도", las=2)
```

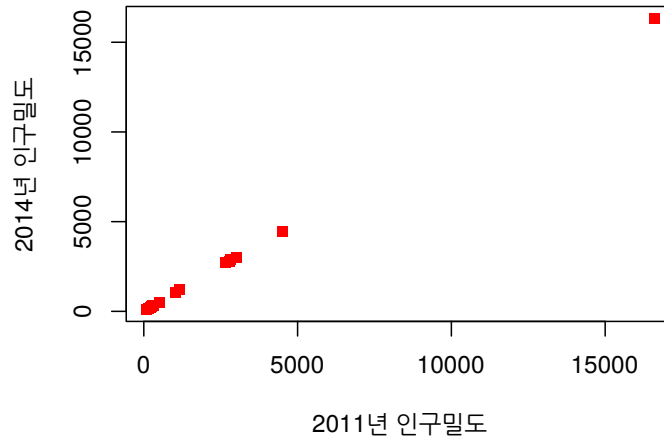
2011년 시도별 인구밀도



6. 위의 자료에서 2011년과 2014년의 지역별 인구밀도를 아래와 같이 그렸다. R코드를 작성하라.

```
plot(x$X2011, x$X2014, xlab="2011년 인구밀도", ylab="2014년 인구밀도",
     pch=15, col="red",
     main="2011년과 2014년의 지역별 인구밀도")
```

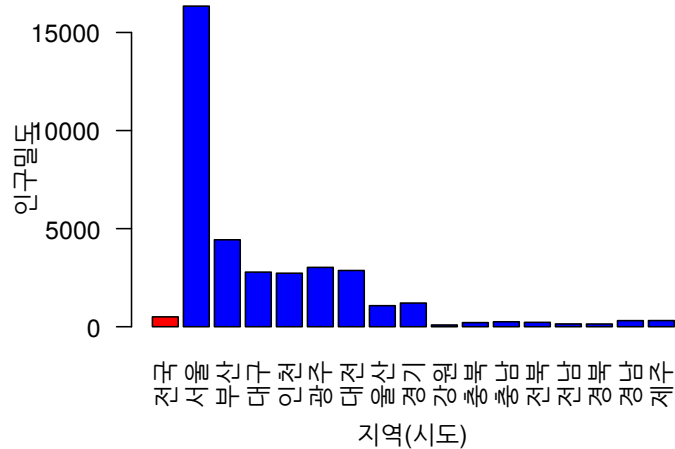
2011년과 2014년의 지역별 인구밀도



7. 위의 자료를 이용하여 2014년 지역별 인구밀도를 아래와 같이 그렸다. R코드를 작성하라.

```
barplot(height = x$X2014, # x$X2014에 대한 막대그래프
        names.arg = x$지역, # x축의 값은 지역(x$지역)
        col=c("red", rep("blue", length(y)-1)), # 막대의 색상은 처음은 빨간색 나머지는 파란색
        ylab="인구밀도", # y축 이름
        xlab="지역(시도)", # x축 이름
        main="2014년 지역별 인구밀도", # 제목
        las=2) # x축의 라벨표시방법(수직으로)
```

2014년 지역별 인구밀도



8. 아래의 R코드의 의미를 상세히 기술하고 결과를 작성하라.

iris데이터프레임을 x로 복사한 후, x의 모든 컬럼들에 대하여 컬럼의 속성(class)가 수치형(numeric)이면 수치형임을 프린트하고 컬럼의 평균과 표준편차를 출력하며, 컬럼속성이 수치형이 아니면 범주형임을 나타내고 범주형 변수의 각 수준별 자료의 수를 출력

```
x <- iris
for(i in 1:ncol(x)){
  if(class(x[,i]) == "numeric"){
    cat(i, "번째 변수는 수치형 변수입니다.\n")
    cat("\t평균은", mean(x[,i]), ", 표준편차는 ", sd(x[,i]), "입니다.\n")
  }else{
    cat(i, "번째 변수는 범주형 변수입니다\n\t각 수준별 자료의 수는 다음과 같습니다.")
    print(table(x[,i]))
  }
}
```

1 번째 변수는 수치형 변수입니다.
평균은 5.843333 , 표준편차는 0.8280661 입니다.

2 번째 변수는 수치형 변수입니다.
평균은 3.057333 , 표준편차는 0.4358663 입니다.

3 번째 변수는 수치형 변수입니다.
평균은 3.758 , 표준편차는 1.765298 입니다.

4 번째 변수는 수치형 변수입니다.
평균은 1.199333 , 표준편차는 0.7622377 입니다.

5 번째 변수는 범주형 변수입니다
각 수준별 자료의 수는 다음과 같습니다.

	setosa	versicolor	virginica
	50	50	50

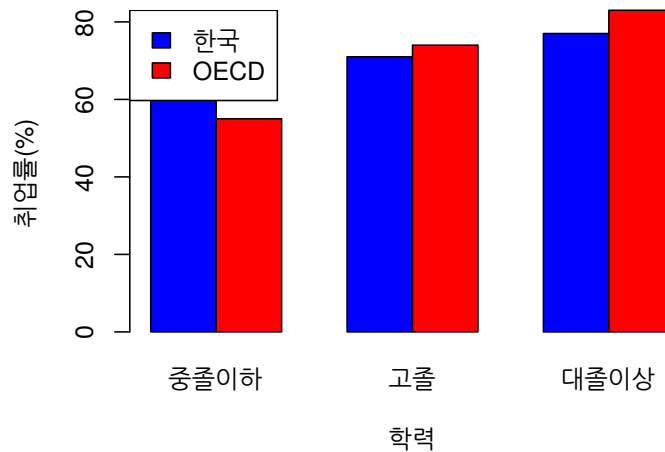
9. 아래의 한국과 OECD국가의 학력별 취업률(%)을 조사한 자료를 이용하여 다음의 막대그림을 그리시오.

	중졸이하	고졸	대졸이상
한국	65	71	77
OECD	55	74	83

```
k <- c(65, 71, 77)
o <- c(55, 74, 83)
x <- rbind(k, o)
rownames(x) <- c("한국", "OECD")
colnames(x) <- c("중졸이하", "고졸", "대졸이상")

barplot(x, beside = TRUE, col = c("blue", "red"), xlab="학력", ylab="취업률(%)")
legend("topleft",rownames(x), fill=c("blue", "red"))
title(main = "한국과 OECD국가의 학력별 취업률 비교", font.main = 4)
```

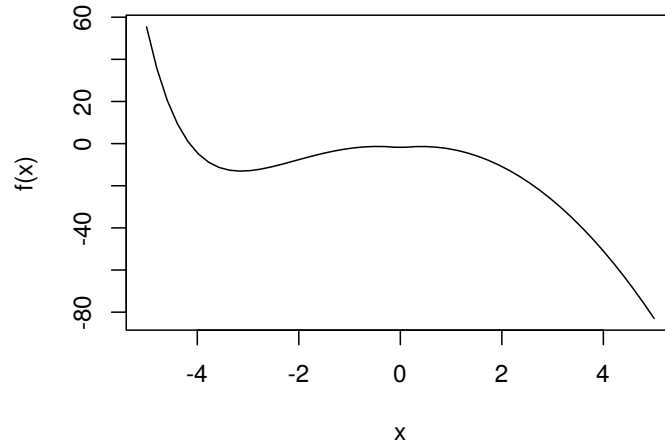
한국과 OECD국가의 학력별 취업률 비교



10. 아래에 정의된 함수 $f(x)$ 에 대하여 주어진 구간($-5 < x < 5$)에서 함수값을 구하는 R함수를 먼저 작성한 후, 그림과 같은 결과를 나타내기 위한 R코드를 작성하고 결과 이미지를 붙여 넣어라.

$$f(x) = e^{-x} - 4x^2 + x + 3|x| - 3$$

```
f <- function(x){
  exp(-x) - 4*x^2 + x + 3*abs(x) - 3
}
x <- seq(-5, 5, length= 50)
plot(x, f(x), type="l")
```



11. 아래 URL의 데이터는 삼성전자 주식의 일별 주가 및 거래량이다.

<http://datamining.dongguk.ac.kr/lectures/2017-1/R/Samsung-stock.txt>

증가 및 거래량을 이용하여 아래와 같이 하나의 그림으로 그리는 R코드를 작성하시오.

```
samsung <- read.csv("http://datamining.dongguk.ac.kr/lectures/2017-1/R/Samsung-stock.txt")
par(mfrow=c(2,1))
plot(samsung$증가, main="삼성전자 증가", type="l", ylab="증가")
plot(samsung$거래량, main="삼성전자 거래량", type="h", ylab="거래량")
```

