R을 이용한 통계 기초(3일차)

인하대학교 대학원 통계학과 국성희

Contents

1. 그래프 그리기
2. 범주형 자료
3. 숫자형 자료
4. 이변량 자료
5. 연습문제

1. 그래프 그리기

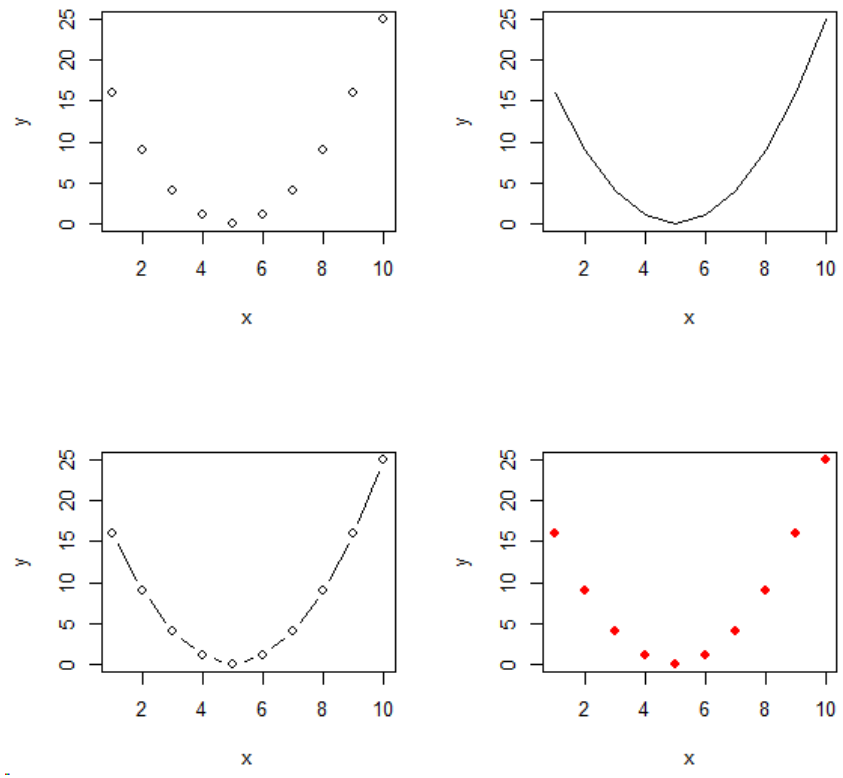
- 1.1 그래프 그리기

```
x=1:10
y=(x-5)^2


plot(x,y)
plot(y~x)
```

1. 그래프 그리기

- 1.1 그래프 그리기

```
#화면 나누기
par(mfrow=c(2,2))

#type별 그래프
plot(x,y,type="p")
plot(x,y,type="l")
plot(x,y,type="b")
plot(x,y,type="p",pch=19,col="red")
```



1. 그래프 그리기

- 1.1 그래프 그리기

  Plot(x,y) 함수에 다음 여러 가지 옵션을 넣어 다양한 형태의 그래프를 그릴 수 있다.

parameter	옵션과 설명
type=	그래프 그릴 때의 형태를 정한다. Type="p" 점(point)으로 그래프를 그린다. type="l" 선(line)으로 그리프를 그린다. type="b" 점과 선으로 이어 그래프를 그린다. type="o" 선이 점 위에 겹쳐진 형태로 그래프를 그린다. type="h" 수직선으로 그래프를 그린다. type="s" 계단(step) 형으로 그래프를 그린다. Type="n" "nothing", 그래프에 아무것도 그리지 않는다.
axes=	T : 디폴트, 축 있게(with axes), F : 축 없게(without axes)
xlim= ylim=	x축과 y축의 상한(upper limit)과 하한(lower limit)를 준다. 예를 들어, xlim=c(1,10) 또는 xlim=range(x)
xlab= ylab=	x축과 y축의 이름을 붙인다.
main=	주요 제목을 그래프 위쪽에 쓴다.
sub=	소제목을 그래프 아래쪽에 쓴다.
pch=	점 표시기호를 선택한다. 예를들어 pch=1 또는 pch="*"
lty=	선으로 그릴 경우 선의 종류를 선택한다. 1 : 실선 2: 파선 3: 점선 ...
col=	색깔을 정한다. "red", "blue", "green" 등을 쓰거나 번호를 준다.

1. 그래프 그리기

- 1.1 그래프 그리기

cars 데이터를 이용하여 산점도에 각 변수의 평균인 x=15.4 , y=42.98와 추정된 직선 y=-17+4x를 그려보자. abline()을 사용한다.

```
data(cars)
attach(cars)
mean(speed)
mean(dist)
```

```
par(mfrow=c(2,2))
plot(speed,dist,pch=1);abline(v=15.4)
plot(speed,dist,pch=2);abline(h=42.98)
plot(speed,dist,pch=3);abline(-17,4)
plot(speed,dist,pch=4)
abline(v=15.4)
abline(h=42.98)
```

1. 그래프 그리기

- 1.1 그래프 그리기

Fiji 섬의 지진 데이터 quakes로 히스토그램 그리기

```
data(quakes)
head(quakes)
par(mfrow=c(1,2))
hist(quakes$mag)
hist(quakes$mag,probability=T,main="histogram with density line")
lines(density(quakes$mag))
```

2. 범주형 자료

- 2.1 데이터의 종류

1. 명목형 자료 : A,B,O,AB형 같은 혈액형
2. 순서형 자료 : A+, A0, B+,B0, C+,C0,….학점
3. 이산형 자료 : 우리나라의 연간 교통사고 건수

4. 연속형 자료 : 시간, 길이, 부피, 넓이 …..

2. 범주형 자료

- 2.1 범주형 자료

```
res=c("y","n","y","y","y","n","n","y","y","y")
table(res)

#막대 그래프
barplot(table(res),xlab="response",ylab="frequency")
barplot(table(res),xlab="response",ylab="frequency",horiz=T)
#파이그림
pie(table(res),main="response“)
```

3. 숫자형 자료

- 3.1 줄기 잎 그림

  관측값의 개수가 너무 많지 않으며 관측값을 줄기와 잎으로 구분할 수 있을 경우 줄기 잎 그림을 그려보면 전체 데이터의 분포를 알 수 있다.

```
#농구경기에서 시합한 팀들의 점수
x=c(45,86,34,98,67,78,56,45,85,75,64,75,75,75,58,45,83,74)
stem(x)
#scale 늘리기
stem(x,scale=2)
```

3. 숫자형 자료

- 3.2 상자그림

  상자그림은 사분위수로 상자를 그리고 최소값과 최대값이 표시되어 데이터의 분포를 대략적으로 알 수 있게 해준다. 이때 사용하는 함수는 boxplot()이다.

```
par(mfrow=c(1,2))
boxplot(x,main="Box Plot", sub="basketball game scores")
boxplot(x,horizontal=T,main="Box Plot", sub="basketball game scores")
```

3. 숫자형 자료

- 3.3 평균, 중앙값, 분산, 표준편자, 사분위수 범위

  (1) 평균 : mean()

  (2) 중앙값 : median()

  (3) 표본 분산 : var()

  (4) 표본 표준편차 : sd()

  (5) 사분위수 범위 : IQR()

  (6) 범위 : range()

  (7) 사분위수 : quantile()

  (8) 기술통계량 요약 : summary()

4. 이변량 데이터

- 4.1 범주형 데이터의 이원분할표

두 개의 범주형 변수가 있는 경우.

부모의 안전벨트 착용여부와 아이의 안전벨트 착용여부를 조사한 빈도 데이터

	아이 안전벨트 착용여부	
부모 안전벨트 착용여부	착용	착용안함
착용	54	7
착용안함	3	12

matrix(c(54,3,7,12),nrow=2) #matrix 함수 이용

rbind(c(54,7),c(3,12)) #rbind 함수 이용

cbind(c(54,3),c(7,12)) #cbind 함수 이용

4. 이변량 데이터

- 4.1 범주형 데이터의 이원분할표

```
#각 행렬에 이름 주기
x= matrix(c(54,3,7,12),nrow=2)
rownames(x)=c("p.buckled","p.unbuckled")
cownames(x)=c("c.buckled","c.unbuckled")
#행과 열의 합
colSums(x)
rowSums(x)
#행렬에 포함
addmargins(x)
```

4. 이변량 데이터

- 4.1 범주형 데이터의 이원분할표

```
#그래프 그리기
par(mfrow=c(1,2))
barplot(x,main="child seat-belt usage")
barplot(x,main="child seat-belt usage",legend.text=T,beside=TRUE)
```

4. 이변량 데이터

- 4.1 범주형 데이터의 이원분할표

#데이터

nicotin	stopsmoke
Y	Y
Y	Y
Y	N
Y	N
Y	Y
N	N
N	N
N	N

왼쪽의 데이터를 csv파일로 저장하고 불러온다.

```
nico=read.csv("파일주소",sep=",",header=T)
attach(nico)
y=table(nicotin,stopsmoke)
prop.table
detach(nico)
```

4. 이변량 데이터

- 4.2 상관계수

machine	expert
68	72
82	84
94	89
106	100
92	97
80	88
76	84
74	70
110	103

왼쪽의 데이터를 txt파일로 저장하고 불러온다.

```
blood=read.table("파일주소",sep=",",header=T)
attach(blood)
cor(machine,expert)
```

5. 연습문제

- 1. 다음은 전구수명을 나타낸 자료이다.

25 16 44 62 36 58 38

(a) 평균 전구 수명을 구하시오

(b) 전구 수명 분산을 구하시오

(c) 전구 수명 표준편차를 구하시오

(d) 상자그림을 그리시오

(e) 줄기잎그림을 그리시오

5. 연습문제

- 2. 내장된 데이터셋인 InsectSprays를 불러들인 후 다음을 구하시오

  (a) spray 종류에 따른 빈도표를 구하시오

  (b) count 평균을 구하시오

  (c) spray 종류에 따른 빈도를 파이그림으로 그리시오

5. 연습문제

- 3. 다음은 지난 1년 동안 A 학과의 각 학년별 학과행사 참석여부를 조사한 데이터다.

	1	2	3	4
참석	40	30	35	20
불참	20	30	45	40

(a) 학년별로 참석 비율을 구하고 막대그래프를 그리시오

(b) 전체적으로 참석 비율을 구하고 막대그래프를 그리시오

Q&A