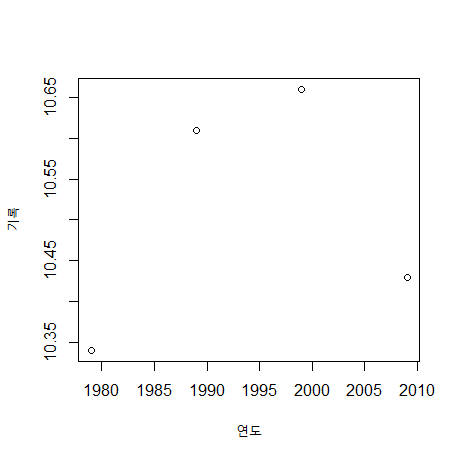
1. 선그림

> plot(exam$기말점수,exam$중간점수)

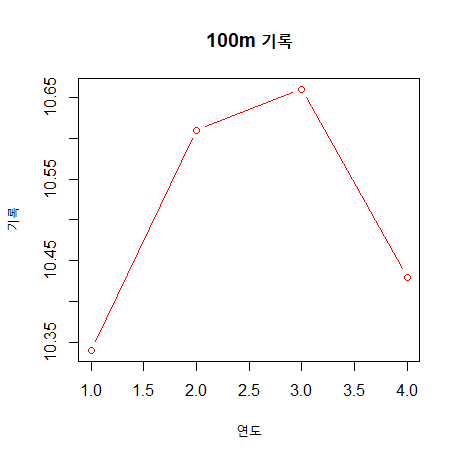
> 기록=c(10.34,10.61,10.66,10.43)

> 연도=c(1979,1989,1999,2009)

> plot(기록~연도)

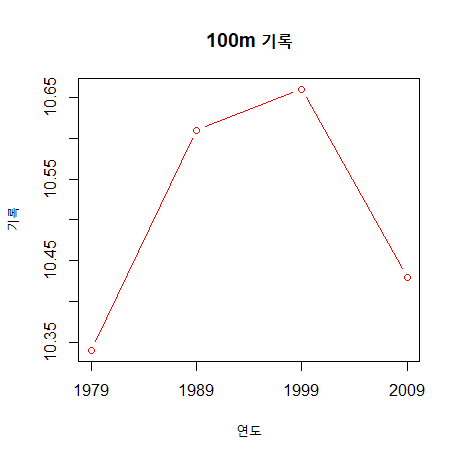


> plot(기록,xlab="연도",type="b",col="red",main="100m 기록")



> plot(기록,xaxt="n",xlab="연도",type="b",col="red",main="100m 기록") // xaxt=”n”을 넣지 않으면 가로축이 겹쳐나옴

> axis(1,at=c(1,2,3,4),label=c("1979","1989","1999","2009"))



2. tapply() 함수

> mg=tapply(exam$중간점수,exam$group,mean)

> mg

1 2 3 4

84.44444 67.45455 77.66667 68.12500 // 4개의 그룹으로 나눠진 수치를 요약하고 싶을 때 사용하는 것, 위의 tapply는 exam안의 4개 그룹안의 자료에 따라 평균을 냄

3. 수치요약

> mean(exam$중간점수,1/10) // 양쪽의 10분의 1씩 절사하고 평균을 구함

[1] 75.9375

> quantile(exam$중간점수) // 4분위수=quantile

0% 25% 50% 75% 100%

7.00 61.50 77.00 87.25 100.00

> quantile(exam$중간점수,c(0.25,0.75))

25% 75%

61.50 87.25

> IQR(exam$중간점수) // IQR=4분위 범위

[1] 25.75

> quantile(exam$중간점수,3/4)-quantile(exam$중간점수,1/4)

75%

25.75

> cor(exam$중간점수,exam$기말점수) // 상관계수 (ex)0.47이면 그렇게 직선에 가깝진 않음)

[1] 0.4764873

// 분산=var(x), 표준편차=sqrt(var(x)) 루트 씌운거임, cov(x,y) 공분산(두개의 퍼짐의 정도를 동시에 알아봄, cor(x,y)=상관계수 -1~1까지 움직임, median(x)=중앙값

4. 두 함수 비교

> fivenum(exam$중간점수)

[1] 7.0 61.0 77.0 87.5 100.0 // 최소값 제1사분위수 중앙값 제3사분위수 최대값

> summary(exam$중간점수)

Min. 1st Qu. Median Mean 3rd Qu. Max.

7.00 61.50 77.00 74.47 87.25 100.00

상자그림은 fivenum임

(1) 홀수 자료: 16, 2, 9, 11 ,4 ,9 ,50, 13, 10 // 자료 개수가 홀수

> x=c(16,2,9,11,4,9,50,13,10)

> fivenum(x)

[1] 2 9 10 13 50

> summary(x)

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.00 9.00 10.00 13.78 13.00 50.00

(2) 짝수 자료: 16, 2, 9, 11, 4, 9, 50, 13 // 자료 개수가 짝수

> y=c(16,2,9,11,4,9,50,13)

> fivenum(y)

[1] 2.0 6.5 10.0 14.5 50.0

> summary(y)

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.00 7.75 10.00 14.25 13.75 50.00

- 통계 프로그램에 따라 4분위수는 조금씩 다르게 나올 수 있음 (어떤 프로그램을 사용했는지 명시해주면 좋음)

5. sample() 함수

> sample(10:50,5) // 10~50까지 숫자 중 5개를 출력

[1] 14 40 30 25 12

> set.seed(3214)

> sample(10:50,5)

[1] 47 48 18 11 34

> set.seed(21)

> sample(10:50,5)

[1] 42 12 25 30 45

> set.seed(21)

> sample(10:50,5)

[1] 42 12 25 30 45 // ex) seed 21에 저장된 랜덤한 5개 숫자출력

마크 맵 시드 같은거임, 시드가 같으면 랜덤 출력이라도 출력값이 같음

6. apply() 함수 // 4주차 2차시 참고

apply(x,1,FUN)

apply(X,2,FUN)

> apply(ch2c,1,sum) // 가운데 1은 행에 대해서 작업할 때

> apply(ch2c,2,sum) // 가운데 2는 열에 대해서 작업할 때

> prop=function(x) x/sum(x) // function 명령어를 통해서 prop이라는 함수를 만듬 (prop=x/[sum(x)])

> apply(ch2c,1,prop) // 만든 함수의 행당 상대도수를 구할 때 씀 (합은 1)

> apply(ch2c,2,prop) // 만든 함수의 열당 상대도수를 구할 때 씀 (합은 1)

> prop.table(ch2c)