

3-2장. 데이터 다루기

박영식(youngsik.park@bsl-lausanne.ch)

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 수학적계산함수

> abs(-5)	# 절대값 계산
[1] 5	
> sqrt(100)	# 제곱근 계산
[1] 10	
> ceiling(3.475)	# 3.475보다 작지 않은 가장 작은 정수
[1] 4	
> floor(3.475)	# 3.475보다 크지 않은 가장 큰 정수
[1] 3	
> trunc(5.99)	# 소수점 이하 버림
[1] 5	
> round(3.475,2)	# 소수 2자리로 반올림
[1] 3.48	
> signif(0.00347, 2)	# 유효수 2자리로 반올림
[1] 0.0035	

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 수학기산함수

> sin(1) [1] 0.841471	# 삼각 함수
> asin(sin(1)) [1] 1	# 역삼각함수
> log(2,base=2) [1] 1	# 밑이 2인 로그
> log(10) [1] 2.302585	# 자연로그
> log10(10) [1] 1	# 상용로그
> exp(log(10)) [1] 10	# 지수함수. 자연로그의 역함수

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 통계계산함수

```
> x<-c(9,8,7,6,5,4,3)
```

```
> mean(x)
```

```
[1] 17.57143
```

```
> median(x)
```

```
[1] 6
```

```
> sd(x)
```

```
[1] 31.98363
```

```
> var(x)
```

```
[1] 1022.952
```

```
> IQR(x)
```

```
[1] 3
```

```
> range(x)
```

```
[1] 3 90
```

```
> sum(x)
```

```
[1] 123
```

```
> min(x)
```

```
[1] 3
```

```
> max(x)
```

```
[1] 90
```

```
[1] 1022.952
```

```
> IQR(x)
```

```
[1] 3
```

```
> diff(c(1,3,6,10,15))
```

```
[1] 2 3 4 5
```

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 문자함수

함수	기능
nchar(x)	문자열 x를 구성하는 문자의 개수
paste(..., sep=" ")	문자열들의 결합
substr(x, start, stop)	문자열의 일부분 선택
toupper(x)	영문자 대문자로 변환
tolower(x)	영문자 소문자로 변환
strsplit(x, split)	문자열의 분리
sub(old, new, x)	문자열의 치환
gsub(old, new, x)	문자열의 치환

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 함수 `nchar()`: 문자열을 구성하고 있는 문자 개수

```
> x <- c("Park", "Lee", "Kim")
```

```
> nchar(x)  
[1] 4 3 3
```

```
> nchar("빅데이터")  
[1] 4
```

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 함수 paste(): 문자열의 결합

▣ 옵션 sep의 활용

```
> paste("모든","사람에게는", " 빅데이터","분석능력이","필요하다")  
[1] "모든 사람에게는 빅데이터 분석능력이 필요하다"
```

```
> paste("모든","사람에게는","빅데이터", " 분석능력이","필요하다", sep="-")  
[1] "모든-사람에게는-빅데이터-분석능력이-필요하다"
```

```
> paste("모든","사람에게는", " 빅데이터", " 분석능력이","필요하다", sep="")  
[1] "모든사람에게는빅데이터분석능력이필요하다"
```

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 입력된 숫자는 문자로 변환됨

```
> paste("원주율은", pi, "이다")  
[1] "원주율은 3.14159265358979 이다"
```

◎ 문자형 벡터가 입력되면 대응되는 요소끼리 결합 벡터에서도 순환법칙이 적용됨

```
> paste("Bigdata",1:3,sep="")  
[1] "Bigdata1" "Bigdata2" "Bigdata3"  
  
> paste(c("Big","data"),1:3,sep="-")  
[1] "Big-1" "data-2" "Big-3"
```


R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 빈칸 없이 문자열 결합

- 1) 함수 paste()에 옵션 sep= "" 사용
- 2) 함수 paste0()에 사용

```
> paste0("stat", 1:3)
[1] "stat1" "stat2" "stat3"
```

◎ 문자형 벡터의 문자열을 하나로 결합: 옵션 collapse

- ▣ 영문 소문자는 letters/ 대문자는 LETTERS: 26개의 영문 대소문자
- ▣ 26개 영문자들을 결합해보자!

```
> letters[1:5]
[1] "a" "b" "c" "d" "e"
> LETTERS[1:5]
[1] "A" "B" "C" "D" "E"
> paste0(letters, collapse="")
[1] "abcdefghijklmnopqrstuvwxyz"
> paste0(LETTERS, collapse=",")
[1] "A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z"
```

26개의 영문 소문자와
대문자를 결합해보세요^^

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 함수 `substr()`: 주어진 문자열의 일부분 선택

`substr(x, start, stop)`

-start, stop: 정수형 스칼라 또는
벡터(대응되는 숫자끼리 시작점과 끝점 구성)

```
> substr("Statistics", 1, 4)
[1] "Stat"
```

```
> x <- c("응용통계학과", "정보통계학과", "학생회장")
```

```
> substr(x, 3, 6)
[1] "통계학과" "통계학과" "회장"
```

```
> substr(x, c(1,3), c(2,6))
[1] "응용" "통계학과" "학생"
```

R 숫자형 벡터와 문자형 벡터를 위한 다양한 함수!

◎ 예제: 문자형 벡터 x에는 미국의 세 도시와 그 도시가 속한 주 이름이 입력

```
> x <- c("New York, NY", "Ann Arbor, MI", "Chicago, IL")
```

◎ 도시가 속한 주 이름만을 선택하여 출력

```
> substr(x, nchar(x)-1, nchar(x))  
[1] "NY" "MI" "IL"
```



주요 이력

現) (주)RTMC 전략기획실장
前) (주)B사 웹로그분석 및 DP사업 完
前) (주)H금속사 회계팀
前) (주)B건설사 회계팀
前) K문고 CRM VIP 군집전략 CRM프로젝트 보조연구원
前) L백화점 CRM Alert 전략 CRM프로젝트 보조연구원

BSL(스위스 로잔 비즈니스 스쿨) MBA
ASSIST 빅데이터경영통계 MBA

국가공인 ADSP(빅데이터 준전문가)

現 코리아IT아카데미 빅데이터 R 강사
現 코리아IT아카데미 빅데이터 기초 파이썬 강사
現 코리아IT아카데미 빅데이터 기초통계 전담강사

“자료는 대가이신 박동련 교수님께 도움을 받았음을 밝힙니다.”

[박영식] [완성에 이르기까지](#)