

# 4-2강

## SAS 함수

Function-name(인수1, 인수2, ...)

- 지정된 인수(argument)를 사용하여 각 함수가 가진 고유의 작업을 수행한 후 결과값을 반환함
- 인자(argument)의 개수는 함수에 따라 정해져 있기도 하고 사용자가 필요한 대로 나열할 수도 있음
- 변수를 쓸 수도 있고 상수를 쓸 수도 있음
- 함수의 종류
  - 수학연산
  - 날짜값, 시간값, 날짜 시간값 처리
  - 요약통계량 계산
  - 문자조정 등 약 150개의 함수가 등록
- SAS도움말과 문서(Document)를 참조

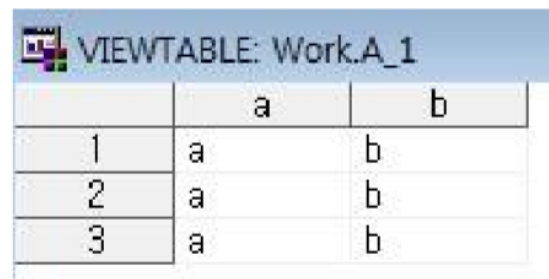
### 문자 함수(1)

함수	정의	예
<code>SUBSTR(arg, start &lt;,length&gt;)</code>	지정한 위치부터 부분 문자열 추출. 길이를 지정하지 않으면 끝까지 추출	<code>SUBSTR('ABCD', 2, 2)</code> → BC
<code>RIGHT(argument)/</code> <code>LEFT(argument)</code>	문자열 오른쪽 정렬 문자열 왼쪽 정렬	<code>RIGHT('AA ')</code> → ' AA' <code>LEFT(' AA')</code> → 'AA '
<code>SCAN(string,n&lt;,delimiters&gt;)</code>	구분자 기준으로 부분 문자열 추출 구분자를 지정하지 않으면 SAS에서 구분자로 정의하는 것 모두 사용	<code>SCAN('A*BC+D', 3, '*+')</code> → D
<code>String1  String2(또는 !!)</code>	함수는 아님. 문자열 결합	<code>'A   '*'  ' C'</code> → A * C
<code>TRIM(argument)</code>	뒤쪽 공백을 지움	<code>TRIM('A ')  '*'  ' C'</code> → A* C
<code>CATX('separator', argument1, argument2,...)</code>	앞뒤 공백제거하고 구분자를 넣어 문자열 결합	<code>CATX('!', 'A ', 'C')</code> → A!C

(예제) SAS trim 옵션과 || 를 이용하여 공백 없이 다른 열의 data 붙이기

- data set 만들기

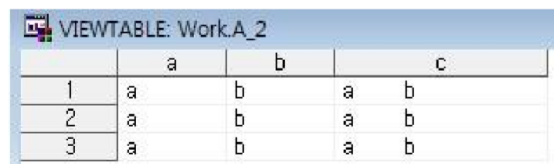
```
DATA a_1;
INPUT a$ b$ @@;
CARDS;
a b
a b
a b
;
RUN;
```



	a	b
1	a	b
2	a	b
3	a	b

- 그냥 || 만으로 붙이기 및 결과

```
DATA a_2;
SET a_1;
c = a||b
RUN;
```



	a	b	c
1	a	b	a b
2	a	b	a b
3	a	b	a b

→ 공백이 그대로 남아 있다.

- trim 명령어를 이용하여 공백 없이 붙이기 및 결과

```
DATA a_3;
SET a_1;
c = trim(a)||b;
RUN;
```

VIEWTABLE: Work.A_3			
	a	b	c
1	a	b	ab
2	a	b	ab
3	a	b	ab

c열이 공백 없이 "a" 와 "b" 가 붙은 것을 확인할 수 있음.

여기서 주의할 점은 || 을 이용하여 붙이기를 할 때, 변수는 문자여야만 한다는 것.

## 문자 함수(2)

함수	정의	예
FIND(target, value <,modifiers,startpos>)	문자열 찾기, 찾은 문자열의 시작 위치를 반환, modifier를 사용하여 대소문자 및 뒤쪽 공백 무시 가능	FIND('STRING', 'in', 'i') → 4 * i : 대소문자 구분없이
INDEX(target, value)	문자열 찾기, 찾은 문자열의 시작 위치를 반환	INDEX('string', 'in') → 4
UPCASE(argument)	대문자로 변환	UPCASE('abc') → ABC
LOWCASE(argument)	소문자로 변환	LOWCASE('ABC') → abc
PROPCASE(string <,delimiters>)	구분자를 중심으로 첫글자는 대문자, 나머지는 소문자로 변환	PROPCASE(john smith, ' ') → John Smith
TRANWRD(source, target, replacement)	특정문자열을 다른 문자열로 치환	TRANWRD('ABC', 'BC', 'bc') → Abc
COMPRESS(var1)	공백 및 특수문자 부분을 삭제	COMPRESS('경상도') → 경상도

→ COMPRESS 정도만 중요하다.

## 문자 함수 SAS 코딩 예1

```

DATA han;
input x $ 10.;
cards;
한국(투자); /*총 10바이트*/
RUN;

DATA han2;
set han;
y=find(x, "(") - 1; /*index도 바이트를 기준으로 나온다. 그러니 한글을 조심할 것*/
z=substr(x,1,y);
RUN;

```

VIEWTABLE: Work.Han2			
	x	y	z
1	한국(투자)	4	한국

## 문자 함수 SAS 코딩 예2

```

DATA substr_test;
text = '010-1234-5678';
var1 = SUBSTR(text,5,4);
var2 = SUBSTR(text,10);/*10부터 다 추출*/
RUN;

```

VIEWTABLE: Work.Substr_test			
	text	var1	var2
1	010-1234-5678	1234	5678

## SAS BASE 시험문제1 (기출)

QUESTION NO: 150

The following SAS program is submitted:

```

data work.test;
Author = 'Christie, Agatha';
First = substr(scan(author,2,' '),1,1);
run;

```

Which one of the following is the value of the variable FIRST in the output data set?

- A. A
- B. C
- C. Agatha

D. " (missing character value)

## SAS BASE 시험문제2(기출)

QUESTION NO: 151

The following SAS program is submitted:

```
data work.test;
  Title = 'A Tale of Two Cities, Charles J. Dickens';
  Word = scan(title,3,' ');
run;
```

Which one of the following is the value of the variable WORD in the output data set?

- A. T
- B. of
- C. Dickens
- D. " (missing character value)

답) B

풀이)

scan(title,3,' ') 는 title에서 구분자를 공백과 콤마로 지정했으니까

A / Tale / of / Two / Cities, / Charles / J. / Dickens

이렇게 나뉘어진다. 여기서 3번째 것을 고르라고 했으니까 of인 B가 정답

## 자주 사용하는 공백제거함수

```
DATA AA;
  INPUT A $4. B $;
  CARDS;

  A 1A B
  A 2A B
  A 3A B;
RUN;
```

	A	B
1	A 1A	B
2	A 2A	B
3	A 3A	B

```
DATA BB;
SET AA;
COMPRESS1=COMPRESS(A);
TRIM2=TRIM('A ')||'B';
CATX3=CATX('Y','A ','B ');
STRIP4=STRIP('ABC DEF ');
RUN;
```

	A	B	COMPRESS1	TRIM2	CATX3	STRIP4
1	A 1A	B	A1A	AB	AYB	ABC DEF
2	A 2A	B	A2A	AB	AYB	ABC DEF
3	A 3A	B	A3A	AB	AYB	ABC DEF

## 수학계산관련함수

함수 형태	내용	예제
<b>ARCOS(argument)</b>	COS의 inverse 값을 계산 $-1 \leq \text{argument} \leq 1$	<b>V=arcos(a);</b> <b>V=arcos(0.3);</b>
<b>COS(argument)</b>	COS 값을 계산 argument은 실수나 Radian값	<b>V=cos(a);</b> <b>V=cos(3.14159/3);</b>
<b>ARSIN(argument)</b>	SIN의 inverse 값을 계산 argument은 실수나 Radian값	<b>V=arsin(a);</b> <b>V=arsin(0.3);</b>
<b>SIN(argument)</b>	SIN 값을 계산 $-1 \leq \text{argument} \leq 1$	<b>V=sin(a);</b> <b>V=sin(3*3.14159);</b>
<b>TAN(argument)</b>	TAN 값을 계산, $-1 \leq \text{argument} \leq 1$	<b>V=tan(a);</b> <b>V=tan(2);</b>
<b>EXP(변수명)</b>	지수함수로 지수 값을 계산한다.	<b>V=EXP(X);</b> <b>V=EXP(3.2);</b>
<b>SQRT(변수명)</b>	제곱근 값을 계산한다. 0 안의 수는 0보다 커야 한다. 제곱은 $x^{**2}$ , 세제곱은 $x^{**3}$	<b>V=SQRT(X); V</b> <b>=SQRT(3.2);</b>
<b>LOG(변수명)</b>	자연 로그(natural log) 값을 계산한다. $\text{Log}_e^x \quad \text{Ln}(X)$	<b>V=LOG(X);</b> <b>V=LOG(3.2);</b>
<b>LOGN(변수명)</b>	로그의 밑이 n인 로그 값을 계산한다. N =10 이면 상용로그 값	<b>V=LOG(X);</b> <b>V=LOG(3.2);</b> <b>V=LOG10(3.2);</b>

## 수학함수 예

```
DATA f2_5;
x1=EXP(1);
x2=GAMMA(4);
x3=LOG(10);
x4=LOG10(100);
RUN;
```

```
PROC PRINT;
RUN;
```

## SAS 시스템

OBS	x1	x2	x3	x4
1	2.71828	6	2.30259	2

## 숫자함수/절단함수

함수	정의	예
SUM(var1, var2, ..., varn)	합을 구함. 결측치를 무시함	SUM(1, 2, .) → 3
MEAN(var1, var2, ..., varn)	평균을 구함. 결측치를 무시함	MEAN(4, 2, .) → 3
ROUND(var1 <,표현할 자리수>)	지정된 자리수에서 반올림(rounding)	ROUND(12.12) → 12 ROUND(-6.47, 0.1) → -6.5
INT(var1)	소수점 아래 절사하고 정수값만 구함	INT(12.12) → 12 INT(-6.67) → -6
CEIL(var1)	올림 정수값만 구함	CEIL(12.12) → 13 CEIL(-6.67) → -6
FLOOR(var1)	내림 정수값 구함	FLOOR(12.12) → 12 FLOOR(-6.67) → -7

## 절단함수 예

```
DATA f2_6;
y1=CEIL(3.58);
y2=FLOOR(-5.4);
y3=INT(-1.58);
y4=ROUND(132.46, 0.1);
RUN;
```

```
PROC PRINT;
RUN;
```

## SAS 시스템

OBS	y1	y2	y3	y4
1	4	-6	-1	132.5

## 날짜 함수

함수	정의	예
TODAY( )	오늘 날짜에 해당하는 SAS 날짜값	오늘이 1960-01-05 이라면, TODAY( ) → 4
MDY(month, day, year)	지정한 월, 일, 년도 값을 이용하여 SAS 날짜값을 생성함	MDY(1, 1, 1960) → 0
YEAR(sas-date)	SAS날짜값에서 연도 추출	YEAR(0) → 1960
MONTH(sas-date)	SAS날짜값에서 월 추출	MONTH(0) → 1
DAY(sas-date)	SAS날짜값에서 일 추출	DAY(0) → 1
QTR(sas-date)	SAS날짜값에서 분기 추출	QTR(0) → 1
WEEKDAY(sas-date)	SAS날짜값에서 요일 추출 (1-일요일 2-월요일...)	WEEKDAY(0) → 6
YRDIF(start,stop,base)	두 SAS날짜간의 연도 차이	YRDIF('16oct1998'd, '16feb2003'd, 'ACT/ACT') ; → 4.3369863014
INTCK(interval,from,to)	두 SAS날짜간의 기간(년,월,일 등등) 차이	INTCK('qtr', '10jan95'd, '01jul95'd) → 2

### • 날짜 함수 예

```
DATA a1;
now=today();
now1=date();
curdate=datetime();
curtime=time();
format now date9. now1 date7. curtime time8.;
RUN;
```

	now	now1	curdate	curtime
1	01JAN2019	01JAN19	1861983219.1	17:33:39

## 형변환 함수



함수	정의	예
INPUT(source, informat)	문자를 숫자로 변환	INPUT('3600', 4.) → 3600 INPUT('3,600', comma5.) → 3600 INPUT('19600101', yymmdd8.) → 0
PUT(source, format)	숫자를 문자로 변환	PUT(3600, 8.) → ' 3600' (문자8byte & 오른쪽 정렬) PUT(0, yymmdd10.) → '1960/01/01' (문자10byte & 오른쪽 정렬)

## 문자변수를 숫자변수로 바꾸기

```
DATA testcn;
INPUT sale $9;
fmtsale = INPUT(sale, comma9.);
CARDS;
2,115,353
;
RUN;
```

OBS	sale	fmtsale
1	2,115,353	2115353

## 숫자변수를 문자변수로 바꾸기

```
DATA test2;
INPUT ymd 8.; /*숫자변수*/
fmtymd = PUT(ymd, $8.); /*문자변수*/
cyear = SUBSTRING(fmtymd, 1, 4);
nyear = INPUT(cyear, 4.); /*문자변수*/
CARDS;
20200305
;
RUN;
```

## 함수의 사용

### 문자 함수 예제

예제 1) 변수타입 변환

- 문자형을 숫자형으로 변환 INPUT (변수, 포맷)
- 숫자형을 문자형으로 변환 PUT (변수, 포맷)

- 데이터 타입 변환은 반드시 다른 변수명을 사용해야 함

```
DATA temp;
x = '12345';
y = INPUT (x, 8.);
z = PUT (y, $8.);
RUN;
```

로그 - (제목없음)

VIEWTABLE: Work.Temp

	x	y	z
1	12345	12345	12345

- 문자형을 숫자형으로 바꾸는 작업은 통계분석을 위해서 매우 중요.  
숫자변수만이 평균, 표준편차 등의 통계량을 구할 수 있기 때문
- 위 예제에서는 모두 12345로 값은 동일하지만 y만이 숫자 변수

#### 예제 2) 주민번호에서 성별 추출

- SUBSTR 함수는 문자열의 일부분을 잘라낼 때 활용
- PUT 문장은 데이터의 중간 계산 과정을 로그 화면에 출력할 때 유용

```
DATA _NULL_; /*_NULL_은 자동변수를 의미한다.*/
id = '801010-1234567';
gender = SUBSTR (id, 8, 1);
PUT id ' -----> ' gender;
RUN;
```

NULL은 work라이브러리에 저장이 안된다.

PUT을 이렇게 사용하면 로그창에 바로 변수에 대한 정보가 나타난다.

로그 - (제목없음)

명령어 ==>

```
146 DATA _NULL_;
147     id = '801010-1234567';
148     gender = SUBSTR (id, 8, 1);
149 PUT id ' -----> ' gender;
150 RUN;
```

801010-1234567 -----> 1

NOTE: DATA 문장 실행:

실행 시간	0.00 초
cpu 시간	0.00 초

## 날짜 함수 예제

예제) 주민번호로 나이 계산

- 주민번호의 연도부분 만을 잘라내어 1900 을 더한다
- 현재날짜의 연도를 계산하여 나이 연수를 만들어 낸다
- 현재날짜는 [DATE()] 또는 [TODAY()]를 사용한다

```
DATA _NULL_;
id = '801010-1234567';
birth_year = 1900 + SUBSTR (id, 1, 2);
today = date();
present_year = YEAR(today);
age = present_year - birth_year + 1;
PUT present_year birth_year age;
RUN;
```

```
1 DATA _NULL_;
2     id = '801010-1234567';
3     birth_year = 1900 + SUBSTR (id, 1, 2);
4     today = date();
5     present_year = YEAR(today);
6     age = present_year - birth_year + 1;
7     PUT present_year birth_year age;
8     RUN;

NOTE: 다음의 위치에서 문자형 값이 숫자형 값으로 변환되었습니다. (행):(칼럼)
3:28
2019 1900 40
NOTE: DATA 문장 실행 (총 프로세스 시간):
실행 시간 0.35 초
cpu 시간 0.06 초
```