

```

/*-----*/
/* 5장 */
/*-----*/
/* 5.1 SAS의 함수 */
/*-----*/
/* CV=변동계수 , NMISS=결측값의 개수, RANGE=범위, STD=표준오차, ABS=절댓값, MOD=나머지, SQRT=제곱근,
ROUND=반올림, INT=정수값*/
/* LENGTH('--') 문자열의 길이 , SUBSTR('--', 시작점, 몇자리) 문자열 나눔*/
/*-----*/

```

```

data Ex5_1;
y1=cv(10,13,14,18);
y2=max(10,13,14,18);
y3=min(10,13,14,18);
y4=mean(10,13,14,18);
y5=n(10,13,14,18);
y6=nmiss(10,.,14,18);
y7=range(10,13,14,18);
y8=std(10,13,14,18);
y9=stderr(10,13,14,18);
y10=sum(10,13,14,18);
y11=var(10,13,14,18);
label
y1='변동계수' y2='최대값' y3='최소값' y4='평균' y5='자료수' y6='결측치 수' y7='범위' y8='표준편차'
y9='표준오차' y10='합계' y11='분산';
run;
proc print label; run;

```

```

/*-----*/

data Ex5_2;
x1=abs(-10);
x2=mod(10,3);
x3=sign(-10);
x4=sqrt(10);
x5=ceil(10.234);
x6=round(10.45,1);
x7=floor(10.434);
x8=int(10.234);
x9=exp(10);
x10=log(10);
x11=log10(10);
x12=cos(3.14/2);
x13=sin(3.14/2);
x14=tan(3.14/2);
label x1='절댓값' x2='나머지' x3='부호' x4='제곱근' x5='올림' x6='반올림' x7='버림' x8='정수'
x9='지수값' x10='자연로그' x11='상용로그' x12='코사인' x13='사인' x14='탄젠트';
proc print label; run;

```

```

/*-----*/
/* SUBSTR 사용할때는 무조건 문자열로, 새로 불러와서 (set) 사용하기 */

data Ex5_3;
input id $ @@;
cards;
20010603 20020131 20021203 20031212 20030701
;
run;

data Ex5_3;
set Ex5_3;
year=substr(id,1,4);
month=substr(id,5,2);
date=substr(id,7,2);
run;
proc print; run;

/*-----*/

data Ex5_4;
input id $ name $ gender $ rep cha min fin;
total = rep+cha+min+fin;
cards;
20193401 SK.KIM M 10 20 40 34
20193405 SI.HWANG M 10 18 38 28
20193410 KS.LEE F 10 19 38 40
20193211 SS.LEE M 10 20 29 34
20193312 MS.SEO F 10 20 38 30
20193115 NR.JUNG F 10 15 20 34
20193120 KW.YEON M 10 19 34 35
20193225 HJ.KIM F 10 20 20 28
20193226 NK.LEE M 10 20 5 10
20193458 JS.BAE M 10 20 38 40
;
run;

data Ex5_4;
set Ex5_4;
dept = substr(id,5,2);
proc format;
    value $deptfmt '31'='MATH' '32'='PHY' '33'='CHEM' '34'='STAT';
run;
proc print data=Ex5_4;
    format dept $deptfmt. ;
run;

/*-----*/
/* 5.2 난수함수 */

```

```

/*-----*/
/* 백분율 및 누적 확률을 보고 싶으면 = freq 프로시저 사용, 균일분포 난수 사용 */

data Ex5_5_1;
do i=1 to 1000 by 1;
    x = ranuni(123)*10;
    y = 1+int(mod(x,6));
    output;
end;
proc freq data=Ex5_5_1;
    table y;
run;

data Ex5_5_2;
do i=1 to 10000 by 1;
    x = ranuni(123)*10;
    y = 1+int(mod(x,6));
    output;
end;
proc freq data=Ex5_5_2;
    table y;
run;

/*-----*/
/* 표준정규분포 난수 사용 */

data Ex5_6;
do i=1 to 10000 by 1;
    x= rannor(123);
    output;
end;
proc means data=Ex5_6 mean std;
    var x;
run;

/*-----*/
/* 구간 확률을 구하려면 if로 베르누이 시행 만들고 freq 프로시저 사용 */

data Ex5_7;
do i=1 to 10000 by 1;
    x= rannor(123);
    if ( -1.96 <= x <= 1.96 ) then y=1;
    else y=0 ;
    output;
end;
proc freq data=Ex5_7;
    table y;
run;

```

```

/*-----*/
/* 이항분포 난수 사용, 표준정규분포의 근사 */

data Ex5_8;
do i=1 to 10000 by 1;
    x = ranbin(123,1200,0.4);
    z = (x-1200*0.4)/sqrt(1200*0.4*0.6);
    output;
end;
proc means data=Ex5_8 mean std;
    var z;
run;

/*-----*/
/* 포아송분포 난수 사용, 포아송분포의 정규근사 */

data Ex5_9;
do i=1 to 1500 by 1;
    x = ranpoi(123,3);
    z = (x-3)/sqrt(3);
    output;
end;
proc means data=Ex5_9;
    var z;
run;

/*-----*/
/* 5.3 확률함수와 분포함수 */
/*-----*/
/* 이산형 분포 */
/* 베르누이분포 ('bern',x,p) 이항분포 ('binom',x,p,n) 초기하분포 ('hyper',x,N,k,n)
음이항분포 ('negb',x-k,p,k) 기하분포 ('geom',x-1,p) 포아송분포 ('poiss',x,m) */
/*-----*/
/* split 띄어쓰기, 음이항, 기하 조심*/

data Ex5_10;
x1 = pdf('binom',4,0.3,10);
x2 = pdf('negb',10-3,0.4,3);
x3 = pdf('poiss',4,3);
x4 = pdf('geom',5-1,0.8);
label
x1='B(10,0.3)일때 * P(X=4)' x2='NB(3,0.4)일때 * P(X=10)' x3='Po(3)일때 * P(X=4)' x4='Ge(0.8)일때 *
P(X=5)';
proc print label split='*'; run;

/*-----*/
/* cdf 범위 계산할때 이산확률변수이므로 뒷부분 -1 해주기 */

data Ex5_11;

```

```

y1=cdf('binom',3,0.5,10)-cdf('binom',1-1,0.5,10);
y2=cdf('negb',8-4,0.8,4)-cdf('negb',6-4-1,0.8,4);
y3=cdf('poiss',3,10)-cdf('poiss',1-1,10);
y4=cdf('geom',3-1,0.8);
y5=cdf('hyper',3,25,10,5)-cdf('hyper',1-1,25,10,5);
label
y1='B(10,0.5)일때 * P(1<=X<=3) ' y2='NB(4,0.8)일때 * P(6<=X<=8) ' y3='Po(10)일때 * P(1<=X<=3) '
y4='Ge(0.8)일때 * P(1<=X<=3) ' y5='HG(25,10,5)일때 * P(1<=X<=3) ';
proc print label split='*'; run;

/*-----*/
/* 연속형 분포 */
/* 균등분포('unif',x,a,b) 정규분포('norm',x,평균,표준편차) 표준정규분포('norm',x,0,1)
지수분포('Expo',x,세타) 베타분포('beta',x,알파,베타) 감마분포('gamma',x,알파,베타)
카이제곱분포('chis',x,df) T분포('t',x,df) F분포('f',x,df1,df2) 로그정규분포('logn',x,평균,표준편차) */
/*-----*/
/* 정규분포에서 표준편차로 입력해야하는것 조심 */

data Ex5_12;
x1 = cdf('norm',135,130,5) - cdf('norm',120,130,5);
x2 = cdf('t',3,10) - cdf('t',1,10);
x3 = cdf('chis',4.5,10) - cdf('chis',3.5,10);
label
x1='N(130,25)일때 * P(120<X<135) ' x2='T(10)일때 * P(1<X<3) ' x3='Chis(10)일때 * P(3.5<X<4.5) ';
proc print label split='*'; run;

/*-----*/

data Ex5_13;
x1=cdf('norm',1,0,1)-cdf('norm',-1,0,1);
x2=cdf('norm',1.64,0,1)-cdf('norm',-1.64,0,1);
x3=cdf('norm',1.96,0,1)-cdf('norm',-1.96,0,1);
x4=cdf('norm',2,0,1)-cdf('norm',-2,0,1);
x5=cdf('norm',2.58,0,1)-cdf('norm',-2.58,0,1);
x6=cdf('norm',3,0,1)-cdf('norm',-3,0,1);
label
x1='P(-1<X<1) ' x2='P(-1.64<X<1.64) ' x3='P(-1.96<X<1.96) ' x4='P(-2<X<2) ' x5='P(-2.58<X<2.58) '
x6='P(-3<X<3) ';
proc print label split='*'; run;

/*-----*/
/* 알파 값 찾기 */
/* 정규분포 Probit(알파) 카이제곱분포 Ciny(알파,df) T분포 (알파,df) F분포 (알파,df1,df2) */
/*-----*/
/* 범위가 반대인 경우는 밖에서가 아니라 안에서 빼준다. */

```

```

data Ex5_14;
x1= probit(0.01);
x2= probit(0.001);

```

```

x3=cinv(0.01,25);
x4=cinv(1-0.05,25);
x5=tinva(0.01,10);
x6=tinva(1-0.025,20);
x7=finv(1-0.05,8,4);
x8=finv(0.1,10,5);

label
x1='P(Z<z)=0.01'    x2='P(Z<z)=0.001'    x3='P(U<u)=0.01'    x4='P(U>u)=0.05'    x5='P(T<t)=0.01'
x6='P(T>t)=0.025'  x7='P(F>f)=0.05'  x8='P(F<f)=0.1';

proc print label ; run;

/*-----*/
/* y축에 표현은 sumvar로 표현, 막대를 나누기 위해서는 discrete 써야한다. */

data Ex5_15_1;
  do x=1 to 10 by 1;
    fx = pdf('binom',x,0.3,10);
    output;
  end;
proc gchart data=Ex5_15_1;
  vbar x /sum sumvar=fx discrete ;
run;

data Ex5_15_2;
  do x=1 to 10 by 1;
    fx = pdf('poiss',x,3);
    output;
  end;
proc gchart data=Ex5_15_2;
  vbar x /sum sumvar=fx discrete ;
run;

/*-----*/
/* gplot 사용해서 정규분포의 그래프 그리기 */

data Ex5_16;
  do x=-3 to 3 by 0.01;
    fx = pdf('norm',x,0,1);
    output;
  end;
proc gplot data=Ex5_16;
  plot fx*x=1;
run;

/*-----*/

```