

001-17  
76125  
L. 80-9

1987-7

# SAS를 이용한 한글처리

1987. 7.

조사통계국 자료관리과

## 일 러 두 기

이 책자는 SAS를 이용하여 한글보고서를 작성할 수 있도록 예제 중심으로 작성한 보고서입니다.

누구나 본서를 이용하여 쉽게 SAS를 이용하여 한글로 결과표를 작성할 수 있도록 편집 하였습니다.

이용중 의문이 있거나 보다 상세한 내용을 아시고자 할 때는 자료관리과 (담당자 송순관)로 연락하여 주시기 바랍니다.

전 화 : 720 - 2788

720 - 2789

1987 년 7 월

자 료 관 리 과 장

## 목 차

가. 개	요 .....	3
나. PROC PRINT	.....	7
다. PROC TABULATE	.....	33
라. PROC SUMMARY	.....	86
마. PROC MEANS	.....	107
바. 한글 자모 영문 Code	.....	116

# 가 . 개 요

1 . 개 요 .....	5
---------------	---

## 1. 개 요

- ① 지금까지 당국에서 Install 한 통계분석 Package 인 SAS 를 사용하는데, 일반 사용자들이 결과표 보고서를 보는데 있어 영문으로 되어 다소 불편한점을 감안 한글 보고서를 만들고자 OUTPUT 내용을 일반 사용자 입장에서 최대한 한글로 바꿀수 있도록 조정하여 좀더 편리하게 SAS 을 이용할 수 있도록 본 program 개발의 목적이 있다.
- ② 지금 SAS 회사에서 개발한 IBM5050 에서는 "EXEC SAS OPTIONS='KANJI'" 로 주어서 직접 OUTPUT 을 화면에서 한글로 볼수 있도록 개발하였지만, 'OUTPUT' 방법이 local printer 로만 한정되어 있어 대량의 data 처리의 output 문제성이 있으므로 당국의 printer 기 상황에 알맞는 IBM 한글을 이용하여 화면에서는 직접 OUTPUT 을 확인할 수 없고 printer 로만 한글 SAS output 보고서를 얻을 수 있도록 하였다.
- ③ 통계분석을 전문으로 하는 전문인의 입장에서 볼 때 실제 전문용어를 한글화 한다는 데는 별 실효성이 없고 보통 사용자들의 입장으로 볼때 일반 reporting 용도인 PRINT, SUMMURY, TABULATE, MEANS 를 중심으로 한글화 OUTPUT 을 얻을 수 있는 예제 중심으로 본서를 꾸며 보았다.
- ④ 통계 분석용 전문 PROC ANOVA, GLM, REG CANCORR ... 등은 실제로 한글로 바꿀수 있는 용어를 본 보고서에서 그 예제를 들지 못했고 FORMAT, LABEL, KEYLABEL, TITLE

## 나. PROC PRINT

1. 개요 .....	9
2. Specifications .....	9
3. PROC PRINT 사용예 .....	10
1) DATA STEP에서 평균, 합계, 분산, 표준편차를 처리하는 방법	10
2) 한글 file을 읽어 처리하는 방법 .....	13
3) Data Step에서 LABEL을 주는 경우 .....	17
4) Variable Length 선언과, PROC STEP에서 LABEL 이용 .....	18
5) NOOBS, DOUBLE, N의 Option을 사용한 예 .....	20
6) PROC에서 BY와 ID를 사용한 예 .....	22
7) PROC에서 SUM을 사용한 예 .....	24
8) Data Step에서 변수를 한글로 바꾸는 과정 .....	26
9) PROC FORMAT을 사용 한글로 변환시키는 방법 .....	28
10) FORMAT을 사용 OUTPUT에 Comma를 주는 방법 .....	31

## 1. 개 요

- 1) PROC PRINT 는 SAS PROC 에서 가장 기본적인 것이므로 SAS data set 이라든가 Disk 에 구성된 file 을 PRINT 할 때 쓰는데 수치변수인 경우는 통계수치와 부분적인 합계를 구할 수 있는 명령을 줄 수 있다.
- 2) 적당한 Option 을 주어서 필요한 보고서를 작성할 수 있다.

## 2. Specification

PROC PRINT Options ;

VAR Variables ;

ID Variables ;

BY Variables ;

PAGEBY Variable ;

SUM Variables ;

SUMBY Variable ;

위의 사용예는 실제 예제를 통해서 설명하였음.

### 3. PROC PRINT 사용예

1) DATA STEP에서 평균, 합계, 분산, 표준편차를 구하는 방법

```

//B19DMNB9 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  ① { LENGTH NAME $14.;
      LENGTH NAM $12.;
  ② INPUT  NAM $ SEX $ X Y Z;
  ③ NAME='?'||NAM||'?';
  ⑤ IF SEX='M' THEN SEX='?SKA WK?';
      ELSE SEX='?DU WK?';
  ⑥ { TOTAL=SUM(X,Y,Z);
      MEAN=MEAN(X,Y,Z);
      STD=STD(X,Y,Z);
      VAR=VAR(X,Y,Z);
  ⑦ FORMAT TOTAL MEAN STD VAR  X Y Z 5.2;
  ④ DROP NAM;
  ⑧ { CARDS;
      RLAGMLDUD F 48 34 36
      GHDGPRUD  F 22 56 46
      QKRGHDUS  F 32 67 78
      RLATJSSU  F 55 87 23
      QKRGMLEHD M 36 90 78
      TLSGHWND  M 77 34 55
      ;
  ⑨ { TITLE1  '?      WHTKXHDRPRNR      TNGKR,DUDDJ,RNRDJWJATN?';
      TITLE2  '-----';
  ⑩ PROC PRINT SPLIT='*' ;
      LABEL
          X  = '?TN  GKR?'
          Y  = '?DUD DJ?'
          Z  = '?RNR  DJ?'
          NAME='?TJD  AUD?'
          SEX = '?TJD  QUF?'
          MEAN=' ?VUD  RBS?'
          TOTAL=' ?CHD  WJA?'
          STD=' ?VUWNS?*?VUS  CK?'
          VAR=' ?QNS  TKS ?';
/*
//

```



- ① NAME 은 Character 로 14 자리, NAM은 12 자리로 선언 하였음. Default값은 8 자리인데 이름을 INPUT 에서 읽어들이는데 8 자리 이상이 있을 것을 가정 12 자리를 NAM으로 선언 하였음
- ② INPUT 하는 Variable 들은 NAM , SEX, X, Y, Z 임
- ③ NAM을 앞뒤로 ?로 접합시켜 한글로 바꾸어 주는 과정을 수행했음. NAME 의 새로운 변수가 생성됨
- ④ NAME 이 생성되었으므로 NAM은 필요없게 되어 DROP시켰음
- ⑤ SEX를 읽어들이는 과정에서 SEX = 'M' 이면 남자로,SEX = 'F' 면 여자로 한글로 바꾸었음
- ⑥ Data Step 에서 통계수치를 구할 수 있음  
X, Y, Z 의 변수를 합하는 key word , 는 SUM(X,Y,Z),  
평균은 MEAN, 편차는 STD, 분산은 VAR의 key word가 있음
- ⑦ TOTAL, MEAN, STD, VAR, X,Y,Z 의 변수의 format 을 5.2 로 주었음 (소숫점 2 자리)
- ⑧ CARDS 로 하고 뒷부분부터 실data 가 주어졌음
- ⑨ TITLE 을 찍어줌 (앞뒤가 ?로 주었으므로 한글로 바뀜)
- ⑩ PROC PRINT 사용예로서 LABEL 은 각변수의 앞뒤를 ?로 접합시켜서 한글로 바꾸어 주었음

\* 결과표는 Table 1

<table 1>

1		자살동기		사망, 언어, 구어적사		9:46 WEDNESDAY, JUNE 10, 1987			
		사망동기		사망, 언어, 구어적사					
095	시 명	시 비	사 하	언 미	구 어	총 점	평 균	비 중	부 사
			하	미	어	점	균	중	사
			하	미	어	점	균	중	사
			하	미	어	점	균	중	사
1	기뢰어	여 자	48.00	34.00	36.00	118	39.33	7.57	57.33
2	중해기	여 자	22.00	56.00	46.00	124	41.33	17.47	305.3
3	바피어	여 자	32.00	57.00	78.00	177	59.00	24.02	577
4	기서너	여 자	55.00	87.00	23.00	165	55.00	32.00	1024
5	바회도	남 자	36.00	90.00	78.00	204	68.00	28.35	804
6	시공	남 자	77.00	34.00	55.00	166	55.33	21.50	462.3

2) 한글 file 을 읽어처리하는 방법

```

//B10DMSSD JOB CLASS=A,MSGCLASS=X
//RUN EXEC HANSAS
① //INDATA DD DSN=A07.MS.LPSMT,DISP=SHR,VOL=SER=BOS241,UNIT=DISK
//SYSIN DD *
  ② DATA DATA1 DATA2;
    ③ KEEP NAME G U L ADD TEL CODE;
      ④ {
        LENGTH NAME $ 32;
        LENGTH G $ 20;
        LENGTH U $ 20;
        LENGTH L $ 20;
        LENGTH ADD $ 8 ;
        LENGTH TEL $ 8 ;
        INFILE INDATA;
        INPUT
          ⑤ {
            @03 CODE $CHAR2.
            @670 NAM $CHAR30.
            @720 G1 $CHAR18.
            @740 U1 $CHAR18.
            @760 L1 $CHAR18.
            @780 ADD $CHAR8.
            @788 TEL $CHAR8. ;
          }
        ⑥ {
          NAME='!||NAM|!';
          U='!||U1|!';
          G='!||G1|!';
          L='!||L1|!';
        }
        ⑦ {
          IF CODE='21' THEN OUTPUT DATA1;
          IF CODE='38' THEN OUTPUT DATA2;
        }
        ⑧ PROC SORT DATA=DATA1 OUT=DATA1;
          BY CODE NAME;
        PROC SORT DATA=DATA2 OUT=DATA2;
          BY CODE NAME;
        ⑨ OPTIONS PAGESIZE=31;
        ⑩ PROC PRINT DATA=DATA1 SPLIT='*';
          VAR NAME G U L ADD TEL;
          ⑪ LABEL NAME='? TKDJQCPAUD?'
            G='? RN TL RNS ?'
            U='? DMQAUSEHD ?'
            L='? QJQWJDFL (EHD)?'
            ADD='? QJSWL GHTN?'
            TEL='? WJSGHK?';
        ⑫ PROC PRINT DATA=DATA2 SPLIT='*';
          VAR NAME G U L ADD TEL;
          ⑬ TITLE '? 1985 SUSEH RUDSKA TKDJQCPAUDQN ?';
          LABEL NAME='? TKDJQCPAUD?'
            G='? RN TL RNS ?'
            U='? DMQAUSEHD ?'
            L='? QJQWJDFL (EHD)?'
            ADD='? QJSWL GHTN?'
            TEL='? WJSGHK?';
        /*
        //
    
```

위의 예제는 광공업 산업편 사업체명과 주소 전화번호를 print 하는 과정을 예를 들었음

- ① 광공업 산업편이 들어있는 data set name은 AO7.MS.LPS-MT 이고 VOL = SER은 BOS241 이고 SAS PROGRAM 에서 INDATA로 읽어 들이겠다고 이름을 줌
- ② 두개의 SAS data set 으로 DATA1, DATA2 로 주었음
- ③ Variable 을 이 data step 에서 유지하고 나머지는 drop 시키겠다는 뜻임
- ④ 각 변수들이 8 자리이상 되는 것은 LENGTH의 명령으로 선언하였음
- ⑤ file을 읽어들이는 명령  
@의 뒤 숫자는 RECORD의 위치를 나타냄  
@ 670 NAM \$ CHAR 30.은 RECORD의 670 column 부터 NAM이라는 변수로 문자값 30 자리를 읽고, 숫자는 \$CHAR 30. 대신 PD30.와 같은 명령으로 읽어들이
- ⑥ INPUT에서 읽어들이는 NAM, U1, G1, L1의 한글 변수를 2byte file code로 file속에 들어있으므로 !로 양끝을 접합시켜 변수를 KEEP(③)에서 준 변수값으로 바꾸어 주었음
- ⑦ 읽어들이는 값중에서 CODE = '21' 이면 DATA1의 SAS data set으로 KEEP해준 변수를 할당하고 CODE='38' 이면 DATA2로 KEEP해준 변수 NAME, G, U, L, ADD, TEL, CODE를 data set에 할당 하게 됨

- ⑧ SORT 문장으로 INPUT data set 를 DATA1 이고 OUT 은 동일 data set 임
- ⑨ PRINT 하는 page 크기는 31 줄로 하고 page 를 넘기겠다는 뜻임. OPTIONS 의 명령으로 줌
- ⑩ PRINT 의 명령으로 data 는 DATA1 이고, 찍고자하는 변수는 VAR NAME G U L ADD TEL 로 나타내었음
- ⑪ 각변수를 한글로 바꾸어 주기 위해서 LABEL 을 주었음  
LABEL 의 명령이 유효하기 위해서는 ⑩에서 LABEL 혹은 SPLIT = '\*' 의 Option 이 필요하다.
- ⑫ 이 PROC 에서는 data set 는 DATA2 로 주었음
- ⑬ TITLE 을 찍어 주었음

\* 결과표는 table 2

< table 2 >

		17:05 MONDAY, JUNE 8, 1987				
1		SAS				
OBS	사업제명	구 시 군	면명도	비지티 (도)	비지 & A	전화
	업 명	구 시 군	명 도	비 명	비 지 & A	화
1						
2	자 가 리 기 여	사 하 구	시 피 도	시 피 도	505	204 2345
3	자 그 도	분 기	하 자 도	하 자 도	271 19	92 2341
4	자 그 오	사 하 구	시 피 도	시 피 도	370 57	203 6101
5	자 대 하 사 기 고 어	분 기	하 자 도	하 자 도	275	93 1960
6	자 대 하 사 사	도래 기	아 라 제 2 도	아 라 도	210	523 0181
7	자 사 모 서	영 더	나 하 도	나 하 도 3 가	109	49 7114
8	자 사 화 사 사 고 자	분 기	자 롱 제 1 도	자 롱 도	692 15	92 1518
9	자 세 위	분 기	하 자 도	하 자 도	400	94 5111
10	자 세 이 사	도래 기	그 사 도	회 위 도	156 1	523 1531
11	자 소 하 사 어	분 기	하 자 도	리 지 도	271 20	324 5161
12	자 지 야 기 기	분 기	하 자 도	하 자 도	234 15	93 6881
13	자 하 도 서 재 사 어	분 기	하 자 도	하 자 도	748 3	92 2526
14	자 지 야 해 배 사 고 자	분 기	가 지 2 도	가 지 도	142 9	92 6911
15	(오) 시 서 재	나 기	대 어 제 3 도	대 어 도	209 32	622 7471
16	(오) 그 저 사 서 투 사 자	도래 기	그 사 도	그 사 도	554 1	57 0552
17	(자) 가 드 그 소	분 기	개 버 도	개 버 도	568 1	92 4224
18	(자) 가 지	분 기	가 지 1 도	가 지 도	123 18	92 9191
19	(자) 거 도	분 기	사 라 도	사 라 도	121 5	326 0618
20	(자) 거 도 사 어	도래 기	영 사 제 9 도	영 사 도	12	85 2574
21	(자) 거 마	분 기	가 지 1 도	가 지 도	955 5	323 5101
22	(자) 거 려 예 보 자	도래 기	그 사 도	회 위 도	153 11	524 3232
23	(자) 거 려 예 지 보	분 기	하 자 도	하 자 도	721 3	92 3133
24	(자) 거 려 예	분 기	가 지 2 도	가 지 도	131 13	92 4500
25	(자) 과 도 시 어	사 하 구	시 피 도	시 피 도	553	29 0117
26	(자) 구 이 사 어 기 기	분 기	가 지 1 도	가 지 도	149 7	323 8612
27	(자) 그 가 트 리	분 기	하 자 도	하 자 도	236 1	93 2017

3) Data Step 에서 LABEL 을 주는 경우

```

//B10DMNBO JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC      HANSAS
//SYSIN    DD *
      DATA DATA1;
        ①LENGTH NAME $14.;
        ②INPUT  NAME $ SEX $ X;
        ③LABEL   X='? DUDDJ?'
                NAME='?TJD AUD?'
                SEX='?TJD QUF?';

        CARDS;
        ?RLAGMLDUD? F 48
        ?GHDGPRUD? F 22
        ?QKRGHUS? F 32
        ?RLATJSSU? F 55
        ?QKRGMLEHD? M 36
        ?TLSGHWND? M 77
        ;
        ④ TITLE1  '?      WHTKXHDRPRNR DUDDJWJATN?';
          TITLE2  '-----';
          PROC PRINT SPLIT='*';

/*
//

```

- ① Variable NAME 은 default 로 8 자리를 읽을 수 있는데 8 자리 이상되는 이름을 읽기 위해서 LENGTH 라는 명령으로 character 로 14 자리를 선언 해 주었음
- ② LABEL 을 DATA step 에서 처리해주는 방법
- ③ TITLE1 은 어느곳에서나 쓰기가 가능한데 뒤의 숫자는 TITLE 이 찍힐 line 의 위치를 설정해 줌
- ④ PROC PRINT 에서 LABEL 문이 유효하기 위해서는 SPLIT 라는 option 이 필요로 함  
SPLIT 대신 LABEL 을 option 으로 줄 수 있다.

\* 결과표는 table 3

< table 3 >

사 사 이 기 기 이 여 여 자 자 사 사 이 기 기 이 여 여 자 자			
-----			
OBS	사 사 이 기	이 기 여 여	자 자
1	기 이 여 이 기	F	48
2	이 기 여 이 기	F	22
3	이 기 여 이 기	F	32
4	이 기 여 이 기	F	55
5	이 기 여 이 기	M	36
6	이 기 여 이 기	M	77

4) Variable Length 선언과 PROC STEP 에서 LABEL 이용

```

//B10DMNB1 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN EXEC HANSAS
//SYSIN DD *
DATA DATA1;
① LENGTH NAME $14.;
① LENGTH NAM $12.;
② INPUT NAM $ SEX $ X;
③ NAME='?'||NAM||'?';
④ IF SEX='M' THEN SEX='?SKA WK?';
ELSE SEX='?DU WK?';
⑤ DROP NAM;
CARDS:
⑥ { RLAGMLDUD F 48
GHDGPRUD F 22
QKRGHDUS F 32
RLATJSSU F 55
QKRGMLEHD M 36
TLSGHWND M 77
;
⑦ TITLE1 '? WHTKXHDRPRNR TNGKRWJATN?';
TITLE2 '-----';
PROC PRINT SPLIT='*';
⑧ LABEL X = '?TN GKR?'
NAME='?TJD AUD?'
SEX = '?TJD QUF?';

/*
//
    
```



- 위 예제에서는 NAM이라는 variable 명으로 읽고 data 앞뒤를 ? 로 접합 시키고 NAM은 필요없으니 DROP시켜 버렸음
- ① LENGTH도 NAME과 NAM을 character로 14자리, 12자리로 선언
  - ② INPUT 시키는 variable
  - ③ NAME을 한글로 바꾸는 step
  - ④ SEX를 Symbol (M, F)에 따라 한글로 바꿈
  - ⑤ 이 data step에서 필요없는 NAM을 delete시킴
  - ⑥ 실Data
  - ⑦ title을 한글로 바꿈
  - ⑧ LABEL을 PROC Step에서 바꾸어줌

\* 결과표는 table 4

<table 4>

조사통계구 사하지스 노 오 기 기 도 F			
OBS	서 미 오 오	시 비 오 노	사 하 F F
1	기희미 오 오	이 자	48
2	호혜기 오 오	이 자	22
3	바리미 기 노	이 자	32
4	기서녀 오 노	이 자	55
5	마희도 기 오	나 자 오	36
6	시희주 노 노	나 자 오	77

5) NOOBS, DOUBLE, N의 Option을 사용한 예

```
//B10DMNB2 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  LENGTH NAME $14.;
  LENGTH NAM $12.;
  INPUT  NAM $ SEX $ X;
  NAME='?'||NAM||'?';
  IF SEX='M' THEN SEX='?SKA WK?';
  ELSE SEX='?DU WK?';

  DROP NAM;
  CARDS;
  RLAGMLDUD F 48
  GHDGPRUD F 22
  QKRGHDS F 32
  RLATJSSU F 55
  QKRGMLEHD M 36
  TLSGHWND M 77
  ;
  TITLE1 '? WTKXHDRPRNR DUDDJWJATN?';
  TITLE2 '-----';
  PROC PRINT NOOBS DOUBLE N SPLIT='*' ;
  LABEL X = '? DUDDJ?'
        NAME='?TJD AUD?'
        SEX = '?TJD QUF?';
/*
//
```

① PROC PRINT STEP에서 Option을 주는 방법

NOOBS : OBS 즉 관측치의 순서를 나타내는 번호가 없음

DOUBLE : Double line으로 결과가 print 됨

N : 관측치의 갯수를 알 수 있음

\* 결과표는 table 5

< table 5 >

자사도계급 소 오 기		여이시사 오 모 T		
서 미 오 오		서 비 오 르		여이 오
기희이 모 오		여 자		48
호혜기 오 오		여 자		22
바하어 기 소		여 자		32
기서녀 모 노		여 자		55
바희도 기 오		나 자 모		36
시후주 노 소		나 자 모		77

6) PROC 에서 BY와 ID를 사용한 예

```

//B10DMNB3 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
      DATA DATA1;
        LENGTH NAME $14.;
        LENGTH NAM  $12.;
        INPUT  NAM $ SEX $ X;
        NAME='?'||NAM||'?';
        IF SEX='M' THEN SEX='?SKA WK?';
        ELSE SEX='?DU WK?';

        DROP NAM;
        CARDS;
        RLAGMLDUD F 48
        GHDGPRUD  M 22
        QKRGHDSU  F 32
        RLATJSSU  F 55
        QKRGMLEHD M 36
        TLSGHWND  M 77
        ;
        TITLE1  '?      WHTKXHDRPRNR  DUDDJWJATN?';
        TITLE2  '-----';
        PROC SORT ;
          BY SEX;
        PROC PRINT SPLIT='*' ;
          ① BY SEX;
            ID NAME;
          ② LABEL
              X  = '? DUDDJ?'
              NAME='?TJD AUD?'
              SEX = '?TJD QUF?';
/*
//

```

위의 예는 PROC PRINT 에서 BY와 ID를 사용하는 예제임  
 BY를 사용하기 이전에는 반드시 PROC SORT 가 선행되어야 한다.  
 다.

- ① BY Variables. 여기서 Variables 은 group data 이어야 함
- ② ID는 관측치를 대표할 수 있는 값

\* 결과표는 table 6

<table 6>

자사통계구 소 오 기		여어지사 오 오 기	
-----		-----	
서 비 오 기		=어 자	
-----		-----	
서 미 오 오		여어 오	
기회어 오 오		48	
바회어 기 소		32	
기서너 오 나		55	
-----		-----	
서 비 오 기		=나 자 오	
-----		-----	
서 미 오 오		여어 오	
호회기 오 오		22	
바회도 기 오		36	
시회주 나 소		77	

7) PROC 에서 SUM을 사용한 예

```

//B10DMNB4 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  LENGTH NAME $14.;
  LENGTH NAM $12.;
  INPUT  NAM $ SEX $ X Y;
  NAME='?'||NAM||'?';
  IF SEX='M' THEN SEX='?SKA WK?';
             ELSE SEX='?DU WK?';

  DROP NAM;
  CARDS;
  RLAGMLDUD F 48 34
  GHDGPRUD  F 22 56
  QKRGHDUS  F 32 67
  RLATJSSU  F 55 87
  QKRGMLEHD M 36 90
  TLSGHWND  M 77 34
  ;
  TITLE1  '?      WHTKXHDRPRNR      TNGKR,DUDDJWJATN?';
  TITLE2  '-----';
  PROC SORT ;
    BY SEX;
  PROC PRINT SPLIT='*' ;
    BY SEX;
    ID NAME;
    ① SUM X Y;
      LABEL
        X  = '?TN GKR?'
        Y  = '?DUDDJ?'
        NAME='? TJD AUD ?'
        SEX = '? TJD QUF ?';
/*
//

```

① 위의 예제에서는 SUM이라는 것을 사용해서 variable X,Y  
의 합계를 성별이 바뀔 때 계산해서 Print 시켰음

\* 결과표는 table 7

< table 7 >

조사대상군		사하, 여미저사	
시 오 기		기 오 기	
-----		-----	
시 비		=여 자	
오 리		오 리	
시 미	사 하	여 미	
오 오	기 기	오	
기희미	48	34	
오 오			
훈해기	22	56	
오 오			
바해미	32	67	
기 사			
기서너	55	87	
오 리			
-----	---	---	
SEX	157	244	
-----		-----	
시 비		=나 자	
오 리		오 리	
시 미	사 하	여 미	
오 오	기 기	오	
바희도	36	90	
기 오			
시해중	77	34	
리 사			
-----	---	---	
SEX	113	124	
	===	===	
	270	368	

8) Data Step 에서 변수를 한글로 바꾸는 과정

```

//B10DMNB5 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
      DATA DATA1;
        LENGTH NAME $14.;
        LENGTH NAM  $12.;
        INPUT  NAM $ SEX $ X Y;
        NAME='?'||NAM||'?';
        IF SEX='M' THEN SEX='?SKA WK?';
           ELSE SEX='?DU WK?';

        DROP NAM;
        CARDS;
        RLAGMLDUD F 48 34
        GHDGPRUD  F 22 56
        QKRGHDUS  F 32 67
        RLATJSSU  F 55 87
        QKRGMLEHD M 36 90
        TLSGHWND  M 77 34
        ;
        TITLE1  '?      WHTKXHDRPRNR      TNGKR,DUDDJWJATN?';
        TITLE2  '-----';
        PROC SORT ;
          BY SEX;
        PROC PRINT SPLIT='*' ;
          BY SEX;
          ID NAME;
          LABEL
            X   = '?TN GKR?'
            Y   = '?DUDDJ?'
            NAME='? TJD AUD  ?'
            SEX = '? TJD QUF  ?';
/*
//

```

\* 결과표는 table 8



<table 8 >

조사대상군		사하, 여미지사	
소 오 기		트 기 오 모 트	
-----		-----	
		시 비	여 자
		오 리	오 자
시 여	사 하	여 미	
오 오	트 기	오	
기 희 여	48	34	
오 오			
호 해 기	22	56	
오 오			
바 히 여	32	67	
기 소 리			
기 서 녀	55	87	
오 리			
-----		-----	
		시 비	나 자
		오 리	오 자
시 여	사 하	여 미	
오 오	트 기	오	
바 희 동	36	90	
기 오			
시 하 주	77	34	
리 소 동			

9) PROC FORMAT을 사용 한글로 변환 시키는 방법

```

//B10DMNB6 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  LENGTH NAME $14.;
  LENGTH NAM $12.;
  INPUT  NAM $ SEX $ DEPAT $ X Y;
  NAM='?'||NAM;
  NAME=NAM||'?';
  IF SEX='M' THEN SEX='?SKA WK?';
              ELSE SEX='?DU WK?';

  DROP NAM;
  CARDS;
  RLAGMLDUD F 1 48 34
  GHDGPRUD  M 2 22 56
  QKRGHDS  F 2 32 67
  RLATJSSU F 1 55 87
  QKRGMLEHD M 1 36 90
  TLSGHWND M 2 77 34
  RLAAUDCJF M 2 34 56
  ;
  TITLE1  '?  WHTKXHDRPRNR  TNGKR,DUDDJWJATN?';
  TITLE2  '-----';
① PROC FORMAT;
  VALUE $SOSOK 1='? WKFYRHK ?'
              2='? QNSTJRRHK?';
② PROC SORT ;
  BY SEX DEPAT;
  PROC PRINT SPLIT='*' ;
  BY SEX DEPAT;
  SUM X Y;
③ PROC FORMAT DEPAT $SOSOK.;
  LABEL
  X  ='? TN GKR?'
  Y  ='? DUDDJ?'
  NAME='? TJD AUD ?'
  SEX ='? TJD QUF ?'
  DEPAT='? THTHRRHK ?';

/*
//

```

- ① 위의 예에서는 FORMAT 를 주는 방법을 제시함  
1 =자료과, 2 =분석과
- ② SORT 는 성별, 소속별로 하였음
- ③ ①에서 바꾸어준 값을 DEPAT 에 주기 위해서 사용한 문장  
\* 결과표는 table 9

< table 9 >

		자사 사 오 기	사하, 여이 사 하, 여이 사 하		
		사 오 기	사 하, 여이 사 하		
-----	서 오 번 리	=여 자	사 오 기	=사 하 과	-----
	OBS	사 오 기	사 하 과	여 이 과	
	1	기 오 희 이 오	48	34	
	2	기 오 서 녀 오 리	55	87	
-----	DEPAT		103	121	-----
-----	서 오 번 리	=여 자	사 오 기	=부 사 하 과	-----
	OBS	사 오 기	사 하 과	여 이 과	
	3	바 오 리 기 사 하	32	67	
-----	SEX		135	188	-----
-----	서 오 번 리	=남 자	사 오 기	=사 하 과	-----
	OBS	사 오 기	사 하 과	여 이 과	
	4	바 오 리 도 기 오	36	90	
-----	서 오 번 리	=남 자	사 오 기	=부 사 하 과	-----
	OBS	사 오 기	사 하 과	여 이 과	
	5	요 오 하 기 오	22	56	
	6	시 오 하 중 리 사 하	77	34	
	7	기 오 미 사 하 리	34	56	
-----	DEPAT		133	146	-----
-----	SEX		169	236	-----
			304	424	

10) FORMAT 을 사용 OUTPUT 에 comma 를 주는 방법

```

//B10DMNB7 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  LENGTH NAME $14.;
  LENGTH NAM $12.;
  INPUT  NAM $ SEX $ DEPAT X ;
  NAME='?'||NAM||'?';
  IF SEX='M' THEN SEX='?SKA WK?';
  ELSE SEX='?DU WK?';

  DROP NAM;
  CARDS;
  RLAGMLDUD F 1 250000
  GHDGPRUD M 1 323450
  QKRGHDUS F 2 260000
  RLATJSSU F 2 345000
  QKRGMLEHD M 2 278000
  TLSGHWND M 2 269000
  RLAAUDCJF M 1 344000
  ;
  TITLE1  '?  WHTKXHDRPRNR  RMQDUDOR  ?';
  TITLE2  '-----';
  PROC FORMAT;
    VALUE  SOSOK 1='?  WKFYRHK  ?'
           2='?  QNSTJRRHK?';
  PROC PRINT SPLIT='*' ;
    SUM X ;
    ① FORMAT DEPAT SOSOK. X COMMA10.;
    LABEL
      X  ='?  RMQDU ?'
      NAME='?  TJD AUD  ?'
      SEX ='?  TJD QUF  ?'
      DEPAT='?  THTHRRHK ?';
/*
//

```

① X 의 값을 summation 하고 각 관측치를 print 할 때 comma를  
 적어주는데 각 수치의 자리는 10 자리를 주었음

\* 결과표는 table 10

<table 10>

OBS	조사대상		급여액	
	서명	성별	직과	급여
1	기희영	여자	자료과	250,000
2	홍희경	여자	자료과	323,450
3	박희영	여자	부서과	260,000
4	기서녀	여자	부서과	345,000
5	박희도	여자	부서과	278,000
6	시희주	여자	부서과	269,000
7	기명자	여자	자료과	344,000
				=====
				2,069,450

## 다. PROC TABULATE

1. 개	요	35
2. Specifications		35
3. 변수를 한글로 바꾸는 방법들		36
1) TITLE을 이용한 방법		36
2) LABEL을 이용한 방법		36
3) KEYLABEL을 이용한 방법		37
4) FORMAT을 이용한 방법		38
4. CLASS 변수를 정하는 방법		40
5. Table 형태를 결정하는 방법		42
6. Option Specifications		44
7. PROC TABULATE Options에서 Option의 사용예		45
1) SAS data set을 선택하는 Option		45
2) MISSING을 선택하는 경우		47
3) table cell의 format을 사용한 예		49
4) ORDER를 사용한 예		51
5) FORMCHAR의 예		54
6) DEPTH의 사용예		57
8. Tabulation 작성 기본 예제		59
1) Table cell의 format 사용예		59
2) 2차원 table의 기본예		62
3) 3차원 table의 기본예		67

4) ALL의 사용예와 LABEL을 직접 바꾸는 방법 .....	72
5) BOX를 한글로 바꾸는 예 .....	77
6) 광공업 산업편 '85년 자료를 이용한 table 예와 page option 을 사용하는 방법 .....	80
7) 산업분류별 종업원 규모별로 생산액 table 예 .....	83
8) 사회통계 시군별, 주택소유구분별 비율(퍼센트) 예 .....	85



## 1. 개요

통계표를 만들 수 있는 PROC으로서 3차원의 표까지 가능함. 즉 페이지별, 행, 열 별로 일반 사용자가 자유로이 표를 구성할 수 있음. 간단한 통계 분석도 가능함. 즉 MEANS, FREQ 그리고 SUMMARY 에서 산출할 수 있는 통계수치를 계산하여 표에 삽입해서 OUTPUT 을 구성할 수 있다.

## 2. Specifications

PROC TABULATE Options ;

CLASS Variables ;

VAR Variables ;

FREQ Variable ;

WEIGHT Variable ;

FORMAT Variables format. ;

LABEL Variable = label ;

BY Variables ;

TABLE {dimension - expression,} {dimension - expression,}  
dimension - expression / options ;

KEYLABEL Keyword = 'text' ... ;

### 3. 변수를 한글로 바꾸는 방법

Tabulation 작성시 모든 변수를 한글로 바꾸어 주는 것이 중요하다. 다른 모든 사항은 보통 SAS 에서 사용하는 것과 동일한데 오직 TITLE, LABEL, KEYLABEL, FORMAT문에서 모든 변수를 한글로 바꾸어 줄 수 있다.

#### 1) TITLE 을 이용한 경우

```
TITLEn n; line 위치를 지정해줌
```

어느 위치에서나 사용 가능하며 다음과 같이 사용할 수 있다.

```
TITLE2 '? GKS RMF ?';
```

위의 예는 2 번째 Line 에 '한글' 이 Print 됨

#### 2) LABEL 을 이용한 경우

보통 input 문에서 생성되는 모든 변수를 다른 이름으로 바꿀 수 있는데 부록에 나와있는 한글 배열로 Key in 하면 된다.

예) DATA DATA1;

```
INPUT NAME $ X Y;
```

```
TOTAL = X+Y;
```

```
LABEL NAME = '? TJDAUD ?'
```

```
X = '? TNGKR ?'
```

```
Y = '? DUD DJ ?'
```

```
TOTAL='? GKQ R& ?';
```

```
CARDS;
```

```
SONG      75      80
```

```

      KIM      60      90
      LEE      75      85 ;

```

위의 예에서 input 문에서 형성된 variable NAME, X, Y, TOTAL 이 각각 성명, 수학, 영어, 합계로 SAS Print 시 한글로 바뀌어진다.

3) KEYLABEL 을 이용한 경우

Tabulation 작성시 생성되는 통계수치의 이름을 일반 사용자가 위의 명령으로 임의로 바꿀 수 있다.

```

예) KEYLABEL  SUM = '? CHD RP ?'
           N    = '? TKDJQCP ?'

```

위의 예에서 SUM, N의 key 통계수치가 “총계, 사업체”로 각각 바뀌었다.

KEYLABEL로 바꾸어 줄 수 있는 통계 변수는 다음과 같다.

N : 관측치 갯수 ( CLASS의 각 분류에 속하는 )

NMISS : 결손치의 갯수

MEAN : 평균

STD : 편차

MIN : 최소값

MAX : 최대값

RANGE : 값의 영역

SUM : 합

USS : 수정하지 않은 제곱합

CSS : 수정한 제곱합

VAR : 분산

STDERR : 평균의 편준 오차

CV :

T :

PRT :

SUM WGT :

PCTN :

PCTSUM :

#### 4) FORMAT

보통 input data 들은 그값을 대표할 수 있는 숫자 ( 1, 2, ..... ),  
혹은 symbol ( A, B, ..... )로 표현해서 data 를 나타낼 수 있다.  
FORMAT 문은 data 의 값을 읽어서 USER 가 알맞는 단어로  
바꾸어 줄 수 있다.

PROC TABULATE 의 문장에서 FORMAT 을 사용하기 이전  
PROC FORMAT 을 사용해서 data 의 값을 바꾸어 주어야 한  
다.

예) DATA SAMPLE ;

```
INPUT NAME $ SEX $ CLASS X Y ;
```

```
CARDS ;
```

SONG	M	1	30	40
KIM	F	1	50	60
LEE	M	2	70	80
PARK	M	2	90	95

```

LU      F      3      80      85
JUNG    F      3      70      75;

```

```
PROC FORMAT;
```

```

VALUE FCLASS 1='? WKFYRHK ?'
                2='? QNSTJRRHK ?'
                3='? TJANRHK ?';

VALUE $SEXX M='? SKAWK ?'
                F='? DUWK ?';

```

위의 예에서는 CLASS의 변수에 읽어 들어 오는 값 1, 2, 3은 각기 FORMAT문에서 1, 2, 3은 각각 “자료과, 분석과, 서무과”로 바뀌어지고 SEX는 M, F가 각각 “남자, 여자”로 바뀌어짐.

주의할 사항은 CLASS의 값은 정수이므로 VALUE FCLASS하고 이름을 주었는데 SEX의 값은 symbol의 값이므로 \$로 시작하는 VALUE \$SEXX로 변수 SEX값을 바꾸어 주었다. TABULATE문에서 위의 바뀐 FORMAT문의 변수를 어떻게 바꿀 것인가 예를 들면

```

PROC TABULATE;

CLASS SEX CLASS;

TABLE SEX, CLASS;

FORMAT SEX $SEXX. CLASS FCLASS;

```

위의 예에서 FORMAT문에 사용한 변수 뒤에 반듯이 “.”가 필요함을 주의할 것

#### 4. CLASS 변수를 정하는 방법

Tabulate Proc 에서 혼동되고 유의할 사항은 variable 를 구분하는데 어느 data 를 CLASS 로 할 것인가는 자주 일어나는 혼동이다. CLASS variable 의 data 는 숫자 ( numeric ) 혹은 문자 ( character ) 일수 있다.

예를들면

```
DATA SAMPLE;  
    INPUT MAME$ SEX $ DEPT MATH ENG;  
    CARDS;  
        ABC    M        1    75    76  
        CBA    F        1    60    65  
        BCA    F        2    80    85  
        CAB    M        2    60    76;
```

위의 예에서 SEX 나 DEPT 는 CLASS 혹은 BY 의 variable 로 사용할 수 있는데 MATH 나 ENG 는 CLASS 혹은 BY 문에서 사용할 수 없는 분석변수 ( Analysis Variable ) 이므로 VAR 에 사용할 수 있다.

통상 CLASS 와 BY 는 같이 쓸 수 없고 분리하여 쓰는 것이 원칙이다.

VAR 에 쓰는 variables 은 반드시 숫자이어야 한다.

위 예의 data set SAMPLE을 가지고 다음을 작성하면

- ① PROC TABULATE DATA = SAMPLE;  
CLASS SEX DEPT;  
VAR MATH ENG;  
TABLE SEX,DEPT\*(MATH ENG);
- ② PROC SORT DATA = SAMPLE;  
BY SEX DEPT;  
PROC TABULATE DATA = SAMPLE;  
BY SEX DEPT;  
VAR MATH ENG;  
TABLE SEX,DEPT\*(MATH ENG)

①에서는 SORT PROC이 필요없고

②에서는 TABULATE사용 이전 BY를 반드시 위해서 반드시 PROC  
SORT가 선행되어야 함

결과표는 ①②와 큰 차이가 있다.

## 5. Table 형태를 결정하는 방법

TABLE 을 작성하는 방법이 좀 다양하게 연습해 둘 필요가 있다. TABLE 에 작성하는 정도에 따라 TABLE 형태가 다양하게 변하기 때문이다.

① 3 차원까지 가능하다 ( page 단위, column, row )

② Table 에 사용하는 operator 의 종류

Comma(,); 새로운 table 의 차원

Asterisk(\*); cross 혹은 subgroup

BLANK ; 접속해서 이어지는 변수 표현

괄호 ( ( ) ) : group 혹은 순서

대괄호 ( < > ) ; 퍼센트를 구할때 분모 명시

등호 ( = ) ; Label 을 바꾸거나 print format 을 정해줌

예) A \* B

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

B \* A

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

A B

A=1	A=2	B=1	B=2	B=3
-----	-----	-----	-----	-----



A \* B C

A=1			A=2			C=1	C=2	C=3
B=1	B=2	B=3	B=1	B=2	B=3			

A \* ( B C )

A = 1						A = 2					
B=1	B=2	B=3	C=1	C=2	C=3	B=1	B=2	B=3	C=1	C=2	C=3

A B \* C

A=1	A=2	B = 1			B = 2			B = 3		
		C=1	C=2	C=3	C=1	C=2	C=3	C=1	C=2	C=3

③ 한 CLASS 의 모든 합을 구할때 사용하는 variable ALL

예) B \* ( A ALL )

B=1			B=2		
A=1	A=2	ALL	A=1	A=2	ALL

( ALL B ) \* ( ALL A )

ALL			B=1			B=2		
ALL	A=1	A=2	ALL	A=1	A=2	ALL	A=1	A=2

## 6. Option Specifications

PROC TABULATE 에 사용되는 Option 들

DATA = SAS dataset

MISSING

FORMAT = format name

ORDER = FREQ

= DATA

= INTERNAL

= FORMATTED

FORMCHAR { index list } = 'string'

DEPTH = number

일반 사용자가 option 에 해당하는 부분을 자유로이 선택해서 원하는 table 결과를 얻을 수 있다.

물론 option 없이 자동처리 ( Default ) 를 함으로써 별다른 어려움 없이 table 의 자료를 볼 수 있는데, option 을 이용해 특수하게 처리해서 원하는 결과표를 얻어보는데 그 의의가 있다. 예를 들면 이미 만들어 놓은 SAS Dataset 이라든가 table 의 cell 의 형식 등이 그 특수한 예이다.

## 7. PROC TABULATE Options 에서 Option의 사용예

### 1) SAS data set 을 선택하는 option

```
//B19DMDDO JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
  ① DATA DATA1;
      INPUT  NAME $ SEX $ AGE HT WT;
      CARDS;
      THDTNSRHKS F 32 163 64
      QKRALGP   F 28 168 54
      GHKDDUDTJR F 27 160 54
      RLATJFGML M 31 160 48
      ;
  ② DATA DATA2;
      INPUT  NAME $ SEX $ AGE HT WT;
      CARDS;
      RLATJFGML M 31 160 48
      QKRALGP   F 28 168 54
      GHKDDUDTJR M 27 160 54
      ;
      PROC FORMAT;
          VALUE $SEX M='?SKAWK?'
                F='?DUWK?';
  ③ PROC TABULATE ;
      CLASS SEX;
      TABLE SEX*F=8.0;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEX. ;
      KEYLABEL N='?AUDTN?';
  ④ PROC TABULATE DATA=DATA1;
      CLASS SEX;
      TABLE SEX*F=8.0;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEX. ;
      KEYLABEL N='?AUDTN?';

/*
//
```

PROC TABULATE DATA=SAS\_data\_set  
option

- 1) DATA STEP 에서 SAS data set 을 만들어 줌  
①②의 DATA1, DATA2
- 2) PROC 에서 data set 을 명시 안하는 경우 가장 최근에 만들어진 SAS data set 이 사용됨  
③의 결과표는 table 1
- 3) PROC 에서 data set 을 명시한 예는 ④의 경우  
결과표는 table 2

< table 1 >

SAS	
-----	
서버	
0 2	
-----	
여자   남자	
0   0	
-----	
미스   미스	
0 1   0 1	
-----	
1   2	
-----	

< table 2 >

SAS	
-----	
서버	
0 2	
-----	
여자   남자	
0   0	
-----	
미스   미스	
0 1   0 1	
-----	
3   1	
-----	

2) MISSING를 선택하는 경우

```

//B19DMDD1 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
      DATA DATA1;
      INPUT  NAME $ SEX $ AGE HT WT;
      CARDS;
      THDTNSRHKS F 32 163 64
①  QKRALGP      . 28 168 54
      GHKDDUDTJR F 27 160 54
      RLATJFGML  M 31 160 48
      ;
      PROC FORMAT;
      VALUE $SEX M='?SKAWK?'
      F='?DUWK?';
②  PROC TABULATE ;
      CLASS SEX;
      TABLE SEX*F=8.0;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEX. ;
      KEYLABEL N='?AUDTN?';
③  PROC TABULATE DATA=DATA1 MISSING;
      CLASS SEX;
      TABLE SEX*F=8.0;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEX. ;
      KEYLABEL N='?AUDTN?';

/*
//

```

PRO TABULATE MISSING  
option

① data card 중 SEX가 결손치 ( missing value ) 임

② option MISSING 없이 사용한 PROC

결과표는 table 3

③ MISSING 사용에 아울러 2 개의 option을 사용하는 방법.

comma 없이 BLANK로 연결함

결과표 table 4

<table 3>

SAS	
서비스 오르	
여자	남자 0
역사 오르	역사 오르
2	1

<table 4>

SAS		
서비스 오르		
	여자	남자 0
역사 오르	역사 오르	역사 오르
1	2	1

3) Table cell 의 format 을 사용한 예

```

//B19DMDD2 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC   HANSAS
//SYSIN    DD *
DATA DATA1;
  INPUT  NAME $ SEX $ AGE HT WT;
  CARDS;
  THDTNSRHKS F 32 163 64
  QKRALGP   F 28 168 54
  GHKDDUDTJR F 27 160 54
  RLATJFGML M 31 160 48
  ;
PROC FORMAT;
  VALUE $SEX M='?SKAWK?'
  F='?DUWK?';
① PROC TABULATE ;
  CLASS SEX;
  TABLE SEX;
  LABEL SEX='?TJDQUF?';
  FORMAT SEX $SEX. ;
  KEYLABEL N='?AUDTN?';
② PROC TABULATE F=7.0;
  CLASS SEX;
  TABLE SEX;
  FORMAT SEX $SEX. ;
  LABEL SEX='?TJDQUF?';
  KEYLABEL N='?AUDTN?';
/*
//

```

PROC TABULATE F=I.D

{ I=정수부분 자리수 }  
{ D=소숫점이하 자리수 }

① Table cell 의 format 없이 PROC 사용

결과표 Table 5 default 값은 12.2 임

② Format 사용예 정수로 7 자리를 출력 함.

결과표 table 6

\* Table 에서 사용한 format 이 option 에서 사용한 format 을  
override 함

<table 5>

SAS	
서버 오르	
여자	남자
역사 오트	역사 오트
3.00	1.00

<table 6>

SAS	
서버 오르	
여자	남자
역사 오트	역사 오트
3	1



4) ORDER 를 사용한 예

```

//B19DMDD3 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC   HANSAS
//SYSIN    DD *
      DATA DATA1;
      INPUT  NAME $ SEX $ AGE HT WT WST;
      CARDS;
      THDTNSRHKS      F    32 163 64 23
      QKRALGP         M    28 168 54 30
      GHKDDUDTJR      M    27 160 54 25
      RLATJFGML       M    31 160 48 24
      ;
      PROC FORMAT;
      VALUE $SEXX F='?DUWK?'
                M='?SKAWK?';
① PROC TABULATE ;
      CLASS SEX;
      TABLE SEX;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEXX. ;
      KEYLABEL N='?AUDTN?';
② PROC TABULATE F=7.0 ORDER=FREQ;
      CLASS SEX;
      TABLE SEX;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEXX. ;
      KEYLABEL N='?AUDTN?';
③ PROC TABULATE F=7.0 ORDER=DATA;
      CLASS SEX;
      TABLE SEX;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEXX. ;
      KEYLABEL N='?AUDTN?';
④ PROC TABULATE F=7.0 ORDER=INTERNAL;
      CLASS SEX;
      TABLE SEX;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEXX. ;
      KEYLABEL N='?AUDTN?';

/*
//

```

```
PROC TABULATE ORDER = FREQ  
                  = DATA  
                  = INTERNAL
```

- ① 컴퓨터 내부에서 생성되는 순서로 table 을 형성함  
④와 같은 결과를 얻음. 결과표는 table 7, table 10
- ② 빈도수가 큰 data 의 항목부터 table 을 형성함.  
결과표는 table 8
- ③ data 가 컴퓨터 내부로 입력되는 순서에 입각해서 table 이  
형성됨. table 9

<table 7 >

서버 오 른	
여자	남자
머스 오 T	머스 오 T
1.00	3.00

<table 8 >

SAS	
서버 오 른	
남자	여자
머스 오 T	머스 오 T
3	1

<table 9 >

SAS	
서버 오 른	
여자	남자
머스 오 T	머스 오 T
1	3

<table 10 >

SAS	
서버 오 른	
여자	남자
머스 오 T	머스 오 T
1	3

5) FORMCHAR 의 예 제

```

//B19DMDD7 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC   HANSAS
//SYSIN    DD   *
DATA DATA1;
  INPUT  NAME $ SEX $ AGE HT WT WST;
  CARDS;
  THDTNSRHKS      F   32 163 64 23
  QKRALGP         F   28 168 54 30
  GHKDDUDTJR     M   31 160 54 25
  RLATJFGML      M   33 160 48 24
  ;
PROC FORMAT;
  VALUE $SEX F='?DUWK?'
           M='?SKAWK?';
① PROC TABULATE FORMCHAR='*****';
  CLASS SEX AGE;
  TABLE SEX*AGE;
  FORMAT SEX $SEX.;
  LABEL SEX='?TJDQUF?'
        AGE='?SKDL?';
  KEYLABEL N='?GKQRP?';
② PROC TABULATE FORMCHAR='';
  CLASS SEX AGE;
  TABLE SEX*AGE;
  FORMAT SEX $SEX.;
  LABEL SEX='?TJDQUF?'
        AGE='?SKDL?';
  KEYLABEL N='?GKQRP?';
③ PROC TABULATE FORMCHAR(1 3 5 9 11)='*****';
  CLASS SEX AGE;
  TABLE SEX*AGE;
  FORMAT SEX $SEX.;
  LABEL SEX='?TJDQUF?'
        AGE='?SKDL?';
  KEYLABEL N='?GKQRP?';
/*
//

```

PROC TABULATE FORMCHAR { index list } = 'string'

① table 의 line 을 ++++ 이러한 문자로 표현한 것임

table 11

② line 을 blank 로 table 12

③ 모서리 부분을 특수 문자로 바꾸는 방법

table 13

< table 11 >

```

SAS
+++++
+               서버               +
+               0 2               +
+++++
+           여자           +           남자           +
+               나이               +           나이               +
+++++
+      28      +      32      +      31      +      33      +
+   하계     +   하계     +   하계     +   하계     +
+   범       +   범       +   범       +   범       +
+++++
+      1.00+   +      1.00+   +      1.00+   +      1.00+   +
+++++

```

< table 12 >

```

SAS
      서버
      0 2
      여자          남자
      나이          나이
      28            32            31            33
      하계         하계         하계         하계
      범           범           범           범
      1.00         1.00         1.00         1.00

```

< table 13 >

```

SAS
*-----*
*               서버               *
*               0 2               *
*-----*
*           여자           *           남자           *
*               나이               *           나이               *
*-----*
*      28      *      32      *      31      *      33      *
*   하계     *   하계     *   하계     *   하계     *
*   범       *   범       *   범       *   범       *
*-----*
*      1.00*   *      1.00*   *      1.00*   *      1.00*   *
*-----*

```

6) DEPTH 의 사용예

```

//B19DMDD4 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC      HANSAS
//SYSIN    DD *
      DATA DATA1;
          INPUT  NAME $ SEX $ AGE HT WT WST;
          CARDS;
          THDTNSRHKS      F   32 163 64 23
          QKRALGP         M   28 168 54 30
          GHKDDUDTJR      M   31 160 54 25
          RLATJFGML       M   31 160 48 24
          ;
          PROC FORMAT;
              VALUE $SEXX F='?DUWK?'
                    M='?SKAWK?';
          ① PROC TABULATE DEPTH=3;
              CLASS AGE;
              VAR HT;
          ② TABLE AGE*HT;
              LABEL AGE='?SKDL?'
                    HT='?TLSWKD?';
              KEYLABEL SUM='?GKQRP?';
/*
//

```

PROC TABULATE DEPTH = number  
 option

① Closing demension 의 차수

ex) A \* B ; 2

② 위의 예에서는 실제로 AGE\*HT\*SUM 의 형식임. table 14

\* default 값은 10 임

통상 default값이 유용히 쓰임.

<table 14>

SAS		
나 이		
28	31	32
시 자 년 오	시 자 년 오	시 자 년 오
하 계 범	하 계 범	하 계 범
168.00	320.00	163.00



## 8. Tabulation 작성 기본 예제

### 1) Table cell 에 format 사용예

```

//B19DMTT1 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC      HANSAS
//SYSIN    DD *
      DATA DATA1;
          INPUT  NAME $ SEX $ AGE HT WT;
          CARDS;
          THDTNSRHKS M 32 163 64
          RLATJFGML F 31 160 48
          QKRALGP F 28 168 54
          GHKDDUDTJR M 27 160 54
          TLSGHWND M 31 165 54
          DLEOGUD M 33 174 64
          AKRUDGML F 32 158 54
          ;
      PROC FORMAT;
          VALUE $SEX M='?SKAWK?'
                F='?DUWK?';
  ① PROC TABULATE;
      CLASS SEX;
      TABLE SEX*N*F=8. ;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEX. ;
      KEYLABEL N='?AUDTN?';
  ② PROC TABULATE;
      CLASS SEX;
      TABLE SEX*F=8.0;
      FORMAT SEX $SEX. ;
      LABEL SEX='?TJDQUF?';
      KEYLABEL N=' ' ;
  ③ PROC TABULATE;
      CLASS SEX;
      TABLE ALL*F=12. SEX*F=12. ;
      LABEL SEX='?TJDQUF?';
      FORMAT SEX $SEX. ;
      KEYLABEL N=' '
                ALL='?GKQR8?';
/*
//

```

① TABLE SEX\*N\*F=8. ;

여기서 N은 sex variable에 속해 있는 group 별 관측치를 알기 위해서 사용함.

F = 8. print 형식을 나타냄 8 자리의 정수.

결과표는 table 1

② ①의 PROC 에서 사용한 것과 동일한 table 형태인데 N과 F = 8.0 이 틀림. ①의 결과와 동일. 틀린 사항은 KEYLABEL 이 틀림. 결과표는 table 2

③ 은 남자 여자의 합계를 알기 위해 key word인 ALL 을 table 에서 사용했음 .

결과표는 table 3

< table 1 >

SAS		
서버 오 른		
여자	남자	
여사 오 른	여사 오 른	
3		4

< table 2 >

SAS		
서버 오 른		
여자	남자	
3		4

< table 3 >

SAS		
서버 오 른		
하계 업	여자	남자
	3	
7		4

2) 2 차원 table 의 기본예

```
//B19DMTT3 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN EXEC HANSAS
//SYSIN DD *
DATA DATA1;
INPUT NAME $ SEX $ SEC AGE HT WT;
CARDS;
THDTNSRHKS M 1 32 163 64
RLATJFGML F 1 31 160 48
QKRALGP F 2 28 168 54
GHKDDUDTJR M 3 27 160 54
TLSGHWND M 3 31 165 54
DLEOGUD M 4 33 174 64
AKRUDGML F 4 32 158 54
RLAWLSQHR F 3 26 157 52
GHDGPRUD F 2 30 154 50
QKRGHDUS F 2 33 150 45
FBWPWJD F 1 29 157 50
DLAUDGH M 1 31 167 56
DLDYDTJR M 4 32 168 62
DHTKARB M 2 30 170 62
QKDDBSGK M 1 38 168 59
GKSTKDRNJS M 4 33 174 65
DLWNSGUD M 1 41 163 60
RLAXOTJR M 2 41 164 59
RLAGUDRNJS M 3 48 167 64
GJSKARJ M 4 38 165 59
TJDDMLTNS F 2 45 157 56
DLRHKSDUD M 1 31 174 75
QKRAUDGML F 3 30 156 54
RLATJSSU F 4 29 155 55
DKSRUDWK F 2 25 160 50
RLAGMLDUD F 2 26 160 50
;
① PROC FORMAT;
VALUE $SEX M='?SKAWK?'
F='?DUWK?' ;
VALUE SECC 1='?QKFRKSR?'
2='?DB ?VKXM?'
3='?WKFYR?'
4='?GIS ?R?' ;
PROC TABULATE;
② CLASS SEX SEC AGE ;
③ TABLE SEX*SEC,AGE*F=8. ;
④ LABEL SEX='?TJDQUF?'
AGE='?DUSFUD?'
SEC='?THHRR?' ;
FORMAT SEX $SEX. SEC SECC.;
KEYLABEL N=' '
ALL='?GKQR?' ;
```

/\*  
//

① FORMAT문에서 SEX는 문자 data이므로 \$로 시작하는 다른 변수를 시작해서 다시 정의해 줌.

SEC은 숫자 data이므로 \$없이 다른 변수로 정의해 줌.

② CLASS 변수는 group화할 수 있는 변수이어야 함.

③ Table이 2차원임

SEX\*SEC는 세로 부분의 계층을 이룸.

, ( comma )는 새로운 차원 즉 가로부분의 새로운 계층.

AGE\*F=8. 은 나이별로 분류를 하되 8자리로 정의해 줌.

결과표는 table 4

④ Table이 2차원임

SEX\*AGE는 남자 여자와 나이로 계층을 이룸.

, ( comma )는 가로부분의 계층.

SEC\*F=12. 은 SEC별로 12자리로 정의해 줌.

결과표는 table 5

< table 4 >

		연령									
		25	26	27	28	29	30	31	32	33	38
여성	사상계										
	바가계	-1	-1	-1	-1	11	-1	11	-1	-1	-1
	DB	11	11	-1	11	-1	11	-1	-1	11	-1
	자계	-1	11	-1	-1	-1	11	-1	-1	-1	-1
	GIS	-1	-1	-1	-1	11	-1	-1	11	-1	-1
남자	바가계	-1	-1	-1	-1	-1	-1	21	11	-1	11
	DB	-1	-1	-1	-1	-1	11	-1	-1	-1	-1
	자계	-1	-1	11	-1	-1	-1	11	-1	-1	-1
	GIS	-1	-1	-1	-1	-1	-1	-1	11	21	11

(CONTINUED)

< table 4 계속 >

SAS

13:45 FRIDAY, JUNE 19, 1987

		여러 L O		
		41	45	48
서버 오 른	스스거 스 거			
여자	바가 거 르 L	.	.	.
	DB 파 트	.	1	.
	자 르 거 르	.	.	.
	GIS 거	.	.	.
남자 오	바가 거 르 L	1	.	.
	DB 파 트	1	.	.
	자 르 거 르	.	.	1
	GIS 거	.	.	.

< table 5 >

SAS 13:45 FRIDAY, JUNE 19, 1987

		사실계 △△			
		바가계 △△	DB 파 △△	자계 △△	GIS 계 △△
서버 이름	이력 년월				
여자	125	.	1	.	.
	126	.	1	1	.
	128	.	1	.	.
	129	1	.	.	1
	130	.	1	1	.
	131	1	.	.	.
	132	.	.	.	1
	133	.	1	.	.
	145	.	1	.	.
남자	127	.	.	1	.
	130	.	1	.	.
	131	2	.	1	.
	132	1	.	.	1
	133	.	.	.	2
	138	1	.	.	1
	141	1	1	.	.
	148	.	.	1	.



3) 3 차원 table 기본예

```

//P19DMTT4 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC  HANSAS
//INDATA   DD   DSN=B19.PROGRAM.LIBRARY(D6DATA),DISP=SHR,UNIT=DISK
//SYSIN    DD   *
      DATA DATA1;
      INFILE INDATA;
      INPUT  NAME $    SEX $    SEC  AGE  HT  WT;
      PROC  FORMAT;
          VALUE $SEX  M='?SKAWK?'
                    F='?DUWK?'  ;
          VALUE  SECC 1='?QKFRKSR&?'
                    2='?DB ?VKXM?'
                    3='?WKFYR&?'
                    4='?GIS ?R&?';
      PROC  TABULATE;
          CLASS SEX SEC AGE ;
          ① TABLE SEC,AGE,SEX*F=12.;
          ② TABLE SEX,AGE,SEC*F=10.0;
          LABEL SEX='?TJDQUF?'
                AGF='?DUSFUD?'
                SEC='?THTHRR&?';
          FORMAT SEX $SEX. SEC SFCC.;
          KEYLABEL N='?AUD?'
                  ALL='?GKQR&?';
/*
//

```

① 3 차원의 예제

SEC 별로 page 가 바뀌고 AGE 는 row, SEX 는 column 으  
로 table 이 형성됨. 결과표는 table 6

② 3 차원 예제

SEX 별로 page 가 바뀜. row 와 column 은 ①과 같음.  
결과표는 table 7

<table 6>

사소계 17	바가계 21	서버 02	
		여자	남자
		0	0
여력 10			
29		1	.
31		1	2
32		.	1
38		.	1
41		.	1

사소계 17	DB 파트 -	SAS	
		서버 02	
		여자	남자
		0	0
여력 10			
25		1	.
26		1	.
28		1	.
30		1	1
33		1	.
41		.	1
45		1	.

< table 6 계속 >

사소계		자라계		SAS	
소기	소기	소기	소기	소기	소기
				서변 오른	
				여자	남자
				명	명
여력 노					
26				1	.
27				.	1
30				1	.
31				.	1
48				.	1

사소계		GIS 계		SAS	
소기	소기	소기	소기	소기	소기
				서변 오른	
				여자	남자
				명	명
여력 노					
29				1	.
32				1	1
33				.	2
38				.	1

< table 7 >

서버 이름	사서계 사기			
	바가계 관리	DB 파트	자라계 관리	GIS계
	명 수	명 수	명 수	명 수
여려 관리				
25	.	1	.	.
26	.	1	1	.
28	.	1	.	.
29	1	.	.	1
30	.	1	1	.
31	1	.	.	.
32	.	.	.	1
33	.	1	.	.
45	.	1	.	.

< table 7 계속 >

서 번 번호	남 자 명	SAS			
		상 계 계			
		바 가 계 계 명	DB 파 트 명	자 라 계 계 명	GIS 계 계 명
		명	명	명	명
27		.	.	1	.
30		.	1	.	.
31		2	.	1	.
32		1	.	.	1
33		.	.	.	2
38		1	.	.	1
41		1	1	.	.
48		.	.	1	.

4) ALL의 사용예와 LABEL을 직접 바꾸는 방법

```

//B19DMTT6 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC   HANSAS
//INDATA   DD DSN=B19.PROGRAM.LIBRARY(D6DATA),DISP=SHR,UNIT=DISK
//SYSIN    DD *
DATA SAMPLE;
  INFILE INDATA;
  INPUT  NAME $      SEX $      SEC AGE HT WT;
  PROC FORMAT;
    VALUE $SEX  M='?SKAWK?'
              F='?DUWK?' ;
    VALUE SECC  1='?QKFRKSR?'
              2='?DB ?VKXM?'
              3='?WKFYR?'
              4='?GIS ?R?' ;
① PROC TABULATE;
  CLASS SEX SEC AGE ;
  VAR HT WT;
  TABLE (ALL SEC) ,(SEX*(WT HT)*SUM)*F=12.;
  LABEL SFX='?TJDQUF?'
        AGE='?DUSFUD?'
        HT  ='?TLSWKD?'
        WT  ='?CPWND?'
        SEC='?THTHRR?' ;
  FORMAT SFX $SEX. SEC SECC.;
  KEYLABEL N= ' '
           SUM='?GKQR?'
           ALL='?GKQR?' ;
② PROC TABULATE ;
  CLASS SEX SEC AGE ;
  TABLE (ALL='?GKQR?' SEC='?THTHRR?')
        (AGE='?DUSFUD?' ALL='?GKQR?')
        (SEX='?TJDQUF?' ALL='?GKQR?')*F=12.;
  FORMAT SEX $SEX. SEC SECC.;
  KEYLABEL N= ' ' ;
/*
//

```

① Table 형태가 2 차원임.

( ALL SEC ) 이므로 합계가 위로 올라 갔음.

결과표는 table 8

② 3 차원의 table로서 특이한 사항은 LABEL 문없이 TABLE  
에서 각변수에 =를 주고 이름을 바꾸어 주었다.

결과표는 TABLE9

< table 8 >

table	SAS			
	시범 영역			
	여자		남자	
	제공 영역	시자 영역	제공 영역	시자 영역
	합계	합계	합계	합계
합계 영역	618	1892	857	2342
소스계 영역				
바가계 영역	98	317	314	835
DB 파트 영역	305	949	121	334
자리계 영역	106	313	172	492
GIS 영역	109	313	250	681

< table 9 >

하계 번호	SAS		하계 번호
	여자	남자	
25	1	.	1
26	2	.	2
27	.	1	1
28	1	.	1
29	2	.	2
30	2	1	3
31	1	3	4
32	1	2	3
33	1	2	3
38	.	2	2
41	.	2	2
45	1	.	1
48	.	1	1
하계 번호	12	14	26



< table 9 계속 >

순서 No	바가계 No	SAS		
		서버 이름		하계 번호
		여자	남자 명	
여력 No				
29		1	.	1
31		1	2	3
32		.	1	1
38		.	1	1
41		.	1	1
하계 번호		2	5	7

순서 No	DB 파트 -	SAS		
		서버 이름		하계 번호
		여자	남자 명	
여력 No				
25		1	.	1
26		1	.	1
28		1	.	1
30		1	1	2
33		1	.	1
41		.	1	1
45		1	.	1
하계 번호		6	2	8

< table 9 계속 >

사수계 표	자르계 표	SAS		
		서벽 오른		하계 표
		여자	남자	
여러 노				
26		1	.	1
27		.	1	1
30		1	.	1
31		.	1	1
48		.	1	1
하계 표		2	3	5

사수계 표	GIS 계	SAS		
		서벽 오른		하계 표
		여자	남자	
여러 노				
29		1	.	1
32		1	1	2
33		.	2	2
38		.	1	1
하계 표		2	4	6

5) BOX 를 한글로 바꾸는 예

```

//P19DMTT5 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC   HANSAS
//INDATA   DD DSN=R19.PROGRAM.LIBRARY(D6DATA),DISP=SHR,UNIT=DISK
//SYSIN    DD *
DATA DATA1;
INFILE INDATA;
INPUT  NAME $      SEX $      SEC  AGE  HT  WT;
PROC FORMAT;
VALUE $SEX  M='?SKAWK?'
            F='?DUWK?' ;
VALUE SFCC  1='?QKFRKSR?'
            2='?DB ?VKXM?'
            3='?WKFYR?'
            4='?GIS?  RB?';

PROC TABULATE;
CLASS SEX SEC AGE ;
① TABLE SEC,AGE,(SEX ALL)*F=12. / BOX='?RN QNS?';
② TABLE (ALL SEC) ,(AGE ALL),(SEX ALL)*F=12. ;
LABEL SEX='?TJDQUF?'
      AGE='?DUSFUD?'
      SEC='?THTHRR?';
FORMAT SEX $SEX. SEC SFCC.;
KEYLABEL N=' '
        ALL='?GKQR?';

```

/\*  
//

① 3 차원의 table 예

특이한 사항은 TABLE문에 스투쉬 ( / ) 뒤에 BOX option 을  
주어서 table box 속에 "구분" ( RN QNS ) 을 넣어 주었음.

결과표는 table 10

② ①의 것과 비슷하나 앞부분에 ALL 이 들어 있으므로 총괄  
표가, 하나 더 Print 됨. Box 의 option 이 없음.

결과표는 table 9

< table 10 >

사실계 사건	바가계 사건	SAS		
		서버 이름		하계 합
가 부 사 건		여 자	남 자	
여러 사 건				
29		1	.	1
31		1	2	3
32		.	1	1
38		.	1	1
41		.	1	1

사실계 사건	DB 파 사건	SAS		
		서버 이름		하계 합
가 부 사 건		여 자	남 자	
여러 사 건				
25		1	.	1
26		1	.	1
28		1	.	1
30		1	1	2
33		1	.	1
41		.	1	1
45		1	.	1

< table 10 계속 >

시소계 구	자리계 구	SAS		
		서비 구		합계
구분 구		여자	남자	
여러 구				
26		1	.	1
27		.	1	1
30		1	.	1
31		.	1	1
48		.	1	1

시소계 구	GIS 구	SAS		
		서비 구		합계
구분 구		여자	남자	
여러 구				
29		1	.	1
32		1	1	2
33		.	2	2
38		.	1	1

6) 광공업 산업별 '85년 자료를 이용한 table 예와 page option을 사용하는 방법

```

//B10DMSAA JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(J,D),REGION=2048K
//RUN EXEC HANSAS
① //INDATA DD DSN=A07.MS.LPPMT,DISP=SHR,VOL=SER=BOS240,UNIT=DISK
//SYSIN DD *
DATA DATA1;
② INFILE INDATA;
INPUT
    @003 SIDO $CHAR2.
    @018 CODE $CHAR8.
    @028 SU PD7.
    @035 MO PD7. ;
    IF '31111>114' <= CODE <= '31119330' THEN OUTPUT;
③ OPTIONS PAGESIZE=60 ;
④ PROC FORMAT;
    VALUE $SIDO0 11='?TJDNF?'
                21='?2NTKS?'
                22='?EORN?'
                23='?DLSCJS?'
                31='?RUDRL?'
                32='?RKDDNJS?'
                33='?CNDQNR?'
                34='?CNDSKA?'
                35='?WJSQNR?'
                36='?WJSSKA?'
                37='?RUDQNR?'
                38='?RUDSKA?'
                39='?WPWN?';
PROC TABULATE F=12.0;
⑤ CLASS CODE SIDO;
⑥ VAR SU MO ;
⑦ TABLE (CODE*(ALL SIDO)) ALL, N (SU MO)*SUM;
    FORMAT SIDO $SIDO0.;
    LABEL CODE='?' 2NSFB '?'
          SIDO='?' TL EH '?'
          SU ='?' TNFID '?'
          MO ='?' RMADOR '?';
    KEYLABEL SUM=' '
          ALL='?SKQRB?'
          N=' ' ;

```

/\*

- ①은 자료가 들어 있는 data set name
- ② ◦ data step으로 SAS data set 은 DATA1  
 INFILE 은 INDATA 여기서 INDATA는 ① // 뒤 INDATA  
 와 동일 이름.
- INPUT 문에서 @003은 record의 위치를 알려 주는 sign.  
 \$CHAR2.는 character로 2자리를 read 하겠다는 의미.
  - @028 SU PD7.은 28 번째 record부터 SU라는 변수로 7자  
 리의 숫자를 읽겠다는 의미.
- ③ print 하는 page 의 line 수를 선택함. 40 줄 찍고 다음 page 로
- ④ SIDO의 변수에 들어오는 값을 한글로 바꾸어 주기 위해  
 서 PROC FORMAT 을 사용했음.
- ⑤ CODE 는 품목, SIDO 는 각시도, 이 두변수는 CLASS 변수  
 로 줄 수 있음
- ⑥ VAR 은 SU는 수량, MO는 금액을 나타냄.
- ⑦ Table에서 세로는 code 별로 구분 하면서 각 code 에 따라 시  
 도로 세분하여, 시도별로 그 합계를 맨처음 ALL  
 로 주어 구하고 마지막 총 code 에 따른 합계를 뒷부분의  
 ALL 로 주어 구하였다. 가로부분의 N은 사업체의 갯수들의  
 미하며 SU MO의 각각 합계는 SUM으로 나타내서 표현하  
 였음.

결과표는 table 11

< table 11 >

SAS				
구분 구분	하계 구분	수량 수	금액 가	금액 가
31119114	시 구	65	16555405	92284
	시 구	37	9742934	56589
	부사 구	2	0	14407
	대 구	1	533173	631
	이 구	1	0	4851
	기 구	11	3468910	5518
	가 구	2	412036	1625
	중 구	2	524177	391
	중 구	3	15175	25
	지 구	1	1859000	2705
	지 구	2	0	5425
	기 구	1	0	0
	기 구	2	0	14
31119119	하계 구분	39	11222911	44412

(CONTINUED)



7) 산업 분류별 종업원 규모별로 생산액 table 예

```

//B10DMSAA JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0),REGION=2048K
//RUN EXEC HANSAS
//INDATA DD DSN=AD7.MS.LPPMT,DISP=SHR,VOL=SER=BOS240,UNIT=DISK
//SYSIN DD *
DATA DATA1;
INFILE INDATA;
INPUT
@018 CODE $CHAR8.
@028 SU PD7.
@035 MO PD7.
@110 SC $CHAR1.;
IF '31119114' <= CODE <= '31119330' THEN OUTPUT;
PROC FORMAT;
VALUE $SCC 1='5-9'
2='10-19'
3='20-49'
4='50-99'
5='100-199'
6='200-299'
7='300-499'
8='500 +';
OPTIONS PAGESIZE=40;
PROC TABULATE F=12.0;
CLASS CODE SC;
VAR SU MO;
TABLE (CODE*(ALL SC)) ALL,N (SU MO)*SUM;
FORMAT SC $SCC;
LABEL SC = '?WHDDNJQDNJS?' ? R3AH?
CODE = '? QNSFB ?'
SU = '? TNFID ?'
MO = '? RMADOR ?';
KEYLABEL SUM = '?GKQR&?'
ALL = '?GKQR&?'
N = '?';

```

결과표는 table 12

< table 12 >

SAS			10:	
구분	종업원 수	인원 수	액 TO	액 TO
31119114	합계	65	16556405	92284
	5-9	4	471327	3400
	10-19	10	772471	1729
	20-49	24	4861555	17440
	50-99	11	7690097	23733
	100-199	12	2760955	37259
	200-299	3	0	0
	300-499	1	0	8723
31119119	합계	39	11222811	44412
	5-9	3	755094	289
	10-19	5	2248469	2221
	20-49	14	2551097	11895
SAS			10:	
구분	종업원 수	인원 수	액 TO	액 TO
31119119	50-99	7	1861222	8836
	100-199	8	2380583	15359
	200-299	1	433600	2194
	500 +	1	987745	3618
31119300	합계	6	319768	794
	10-19	2	39779	189
	20-49	2	1100	60
	100-199	1	106619	365
	200-299	1	172270	179
합계		110	23098984	137490

8) 사회통계 시군별 주택소유 구분별 비율 ( Percent 계산 ) 예

```

//E10DMSIC JOB CLASS=A,MSGCLASS=X
//STEP1 EXEC HANSAS
//R DD DSN=A20.S1.YSM,UNIT=DISK,DISP=SHR,VOL=SER=60S242
//SYSIN DD *
DATA A ;
KEEP A B D1-D4 GU COUNT;
INFILE R ;
INFUT
      A 13
      D1 14-16
      D2 17-19
      D3 20-22
      D4 23-25
      GU 53 ;
COUNT=1;
IF GU=1 OR GU=2 THEN B=1;
ELSE B=2;
PROC FORMAT;
VALUE GUU 1='?EOEHTL?'
          2='?THEHTL?'
          3='?RNSQN?';
VALUE AA 1='?TKSEK?'
          2='?DE?'
          3='(?DB?)'
          4='?AN?';
VALUE BB 1='?EHTL?'
          2='?RNSQN?';
OPTIONS PAGESIZE=45;
PROC TABULATE DATA=A F=11.;
CLASS GU A B ;
VAR COUNT D1 D2 D3 D4 ;
TABLE (ALL='?CHDRKNTN?' ALL='?WJSRNR?'*(A*PCTN<A*COUNT>))
      (B*(ALL='?GKQR?' A*PCTN<A*COUNT>))
      E*(GU*(A*PCTN<E*A*COUNT> ALL='?THR?')),COUNT D1 D2 D3/RTS=44;
FORMAT GU GUU. A AA. B BB. ;
LABEL COUNT='?VJTPSXM?'
      GU='?RNSQN?'
      A = '?THDBGUF?'
      B = '?TLRNSQN?'
      D1= '?QNAH?'
      D2= '?WJSTP?'
      D3= '?WJCNR?'
      D4= '?DBDWK?';
KEYLABEL SUM=' '
      PCTN=' ' ;
/*
//

```

\* 결과표는 table 13

< table 13 >

		SAS	10:42 TUESDAY, JULY 28, 1987		
		피사열 L -	번호 TA	지사열 L	지점 기
총계 O TT		17660	113340	76437	318626
지점 L 기	ACOM AVE				
	사다 L	46	9	11	24
	O W	4	1	1	2
	( O ) W	7	2	2	4
	O T	42	0	0	0
시점 LT	시점 B	11529	59600	70313	198922
다시 A	ACOM AVE				
	사다 L	35	8	16	24
	O W	4	1	1	3
	( O ) W	9	2	3	5
	O T	52	0	0	0
계 LT	시점 B	6131	53740	6124	119704
	ACOM AVE				
	사다 L	69	12	2	25
	O W	3	1	1	2
	( O ) W	4	1	0	2

< table 13 계속 >

			SAS	10:42 TUESDAY, JULY 28, 1987		
			표시 L -	부 T A	지시 L	지시 T
시 L T		사 A O M Y				
계 L T		0 T	24	0	.	.
시 L T	계 T L	사 A O M Y				
다 A	다 A A	사 A 다 L	34	7	19	24
		0 T	4	1	1	3
		( 0 ) T	9	2	4	6
		0 T	53	0	0	0
		사 A 계 L T	7480	37733	52982	128801
	다 A A	사 A 다 L				
		사 A 다 L	36	8	11	22
		0 T	5	1	1	3
		( 0 ) T	9	2	2	5
		0 T	50	.	0	0
		사 A 계 L T	4049	21867	17331	70121
계 L T	계 L T	사 A 다 L				
		사 A 다 L	69	12	2	25
		0 T	3	1	1	2

< table 13 계속 >

				SAS 10:42 TUESDAY, JULY 28, 1987			
			표시 L -	부 T A	지시 L	지시 L	
시 L T	부 T L	부 T A					
부 T L	부 T L	( 0 )	4	1	0	2	
		0	24	0	.	.	
		부 T A	6131	53740	6124	119704	

## 라. PROC SUMMARY

1. 개    요 .....	91
2. Specifications .....	91
3. OUTPUT 에 형성될 수 있는 통계결과 .....	92
4. PROC SUMMARY 예제 .....	93
1) OUTPUT 에서 SAS data set 을 주는 방법 .....	93
2) 결과표를 보는 방법 .....	95
3) 평균, 합계를 구하는 예 .....	97
4) 분산을 구하는 경우 .....	102
5) 변수를 선택해서 통계결과를 구하는 예 .....	103
6) 가중치를 주어서 통계결과를 구하는 예 .....	105

## 1. 개 요

PROC SUMMARY를 이용해서 기술통계 결과 보고서 작성에 편리한 PROC 임.

특이할 사항은 Print output이 생성되지 않으므로 output SAS data set인 MYDATA를 반드시 기술해 주어야 하고, PROC PRINT를 사용해서 결과표를 얻어야 한다.

## 2. Specifications

```
PROC SUMMARY option ;  
    CLASS variables ;  
    VAR variables ;  
    BY variables ;  
    FREQ variable ;  
    WEIGHT variable ;  
    ID variables ;  
    OUTPUT OUT = MYDATA
```



### 3. OUTPUT 에 형성될 수 있는 통계결과

N ; 그룹별 관측치 갯수

NMISS ; 그룹별 결손치 갯수

MEAN ; 평균

STD ; 표준편차

MIN ; 최소값

MAX ; 최대값

RANGE : 범위

SUM ; 합계

VAR ; 분산

USS ; 수정되지 않은 제곱합

CSS ; 수정된 제곱합

CV

STDERR

T

PRT

SUMWGT 등.

#### 4. PROC SUMMARY 예제

##### 1) OUTPUT 에서 SAS dataset 을 주는 방법

```

//B19DMTTO JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  ① INPUT  NAME $ JUN FEB MAR;
     CARDS;
     A 12000  32000  34000
     C 32000  23450  32110
     B 23350  22450  12900
     B 32000  34890  12349
     C 21280  21340  12359
     A 32870  23345  12345
     ;
  ② PROC FORMAT;
     VALUE $COMPANY A='?GUSEO?'
                  B='?RMREHD?'
                  C='?EHDDK?';
  ③ PROC SUMMARY DATA=DATA1;
  ④ CLASS NAME;
  ⑤ VAR JUN FEB MAR;
  ⑥ OUTPUT OUT=MYDATA;
  ⑦ PROC PRINT SPLIT='*' DATA=MYDATA;
     TITLE1  '?      WPWHDJQRP 1DNJF - 3DNJF AOCNFDOR?';
     TITLE2  '-----';
  ⑧ FORMAT NAME $COMPANY.;
  ⑨ LABEL  NAME='?GHLTKAUD?'
           _TYPE_='?GUDXO?'
           _FREQ_='?QLSEII?' ;
/*
//

```

- ① DATA STEP으로 DATA1 의 SAS data set 을 만들어 주고 variable 은 NHME JUN FEB MAR 로 data 를 읽어 들임.
- ② FORMAT문으로서 A, B, C가 문자이므로 VALUE \$ COMPANYY 로 바꾸어 줌.

- ③ PROC SUMMARY 이고 SAS data set 은 DATA1임.
- ④ CLASS data는 group이 가능한 data 이어야 함.
- ⑤ 분석 가능한 수치 data variable 들
- ⑥ OUTPUT 은 OUT= MYDATA 를 반드시 기술하고
- ⑦ PRINT STEP 에서 ⑥에서 형성한 SAS data set 인 MYDATA 를 기술해 주어야 함
- ⑧ 변수 NAME 을 ②에서 형성한 \$COMPANY. 로 바꾸어 주는 문장
- ⑨ LABEL 문으로 NAME 은 이름으로 =FREQ=, -TYPE= 은 PROC SUMMARY 에서 형성되는 자동적인 값임.

\* 결과표는 table 1

<table 1 >

제지어계 1위 - 3위 매출액			
-----			
OBS	회사명	형태	빈도
① 1		0	6
2	현대	1	2
3	그도	1	2
4	도아	1	2

위의 예에서

①은 OBS 1로서 형태는 0으로 최상위 level 을 나타내고, 빈도는 6으로 총 data 수를 의미하고, 형태 1은 subgroup 으로 각회사가 2개씩 있다는 의미임.

2) 결과표를 보는 방법

```

//B19DMTT1 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  INPUT  NAME $ REGION JUN FEB MAR;
  CARDS;
  A 2 12000 32000 34000
  A 1 12340 32230 34450
  C 1 32000 23450 32110
  C 2 32230 23457 32111
  C 1 32222 22222 32122
  B 1 23350 22450 12900
  B 1 24440 22345 12912
  B 2 32000 34890 12349
  C 2 21280 21340 12359
  A 1 32870 23345 12345
  ;
PROC FORMAT;
  VALUE $COMPANY A='?GUSEO?'
                B='?RMREHD?'
                C='?EHDDK?';
  VALUE REGIONN 1='?QN TKS?'
                2='?TJDNF?';
PROC SUMMARY DATA=DATA1;
  CLASS NAME REGION;
  OUTPUT OUT=MYDATA;
PROC PRINT SPLIT='*' DATA=MYDATA;
  TITLE1 '? WPWHDJQRP 1DNJF - 3DNJF AOCNFDOR?';
  TITLE2 '-----';
  FORMAT NAME $COMPANY. REGION REGIONN.;
  LABEL NAME='?GHLTKAUD?'
        REGION='?WLDURQUF?'
        _TYPE_='?GUDXO?'
        _FREQ_='?QLSEH?' ;
/*
//

```

앞의 예제와 유사한 경우인데 CLASS data가 NAME 과 REGION으로 구성되어 있음.

\* 결과표는 table 2

<table 2 >

지역이계 1위 - 3위 매출액				
OBS	회사명	지역별	형태	빈도
1		.	0	10
2		부산	1	6
3		서울	1	4
4	현대	.	2	3
5	그도	.	2	3
6	도아	.	2	4
7	현대	부산	3	2
8	현대	서울	3	1
9	그도	부산	3	2
10	그도	서울	3	1
11	도아	부산	3	2
12	도아	서울	3	2

OBS 1은 빈도가 10 개이므로 총 회사수를 의미하고 OBS 2, 3은 부산에 회사가 6 서울에 4 개를 의미하며 형태가 1인 Level 로 구성되었음. OBS 4~6 회사명대로 회사의 갯수를 의미하며 형태는 2 Level 로 이루어졌음.

OBS 7~12는 회사별로 지역에 몇개씩 분포되어 있나를 나타냄. 형태는 3 Level 로 이루어졌음.

2) 결과표를 보는 방법

```

//B19DMTT1 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  INPUT  NAME $ REGION JUN FEB MAR;
  CARDS;
  A 2 12000 32000 34000
  A 1 12340 32230 34450
  C 1 32000 23450 32110
  C 2 32230 23457 32111
  C 1 32222 22222 32122
  B 1 23350 22450 12900
  B 1 24440 22345 12912
  B 2 32000 34890 12349
  C 2 21280 21340 12359
  A 1 32870 23345 12345
  ;
PROC FORMAT;
  VALUE $COMPANY A='?GUSEO?'
                B='?RMREHD?'
                C='?EHDDK?';
  VALUE REGIONN 1='?QN TKS?'
                2='?TJDNF?';
PROC SUMMARY DATA=DATA1;
  CLASS NAME REGION;
  OUTPUT OUT=MYDATA;
PROC PRINT SPLIT='*' DATA=MYDATA;
  TITLE1 '? WPWHDJQRP 1DNJF - 3DNJF AOCNFDOR?';
  TITLE2 '-----';
  FORMAT NAME $COMPANY. REGION REGIONN.;
  LABEL NAME='?GHLTKAUD?'
        REGION='?WLDURQUF?'
        _TYPE_='?GUDXO?'
        _FREQ_='?QLSEH?' ;
/*
//

```

앞의 예제와 유사한 경우인데 CLASS data가 NAME 과 REGION으로 구성되어 있음.

\* 결과표는 table 2

<table 2>

지역이계 1위 - 3위 매출액				
OBS	회사명	지역별	형태	빈도
1			0	10
2		부산	1	6
3		서울	1	4
4	현대		2	3
5	그도		2	3
6	도아		2	4
7	현대	부산	3	2
8	현대	서울	3	1
9	그도	부산	3	2
10	그도	서울	3	1
11	도아	부산	3	2
12	도아	서울	3	2

OBS 1은 빈도가 10개이므로 총 회사수를 의미하고 OBS 2, 3은 부산에 회사가 6 서울에 4개를 의미하며 형태가 1인 Level로 구성되었음. OBS 4~6 회사명대로 회사의 갯수를 의미하며 형태는 2 Level로 이루어졌음.

OBS 7~12는 회사별로 지역에 몇개씩 분포되어 있나를 나타냄. 형태는 3 Level로 이루어졌음.

3) 평균, 합계를 구하는 예

```

//B19DMTT5 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC      HANSAS
①//INDATA DD DSN=B19.PROGRAM.LIBRARY(D6DATA),DISP=SHR,UNIT=DISK
//SYSIN   DD *
  DATA DATA1;
  ② { INFILE INDATA;
      INPUT  NAME $   SEX $   SEC AGE HT WT;
      ③ PROC FORMAT;
          VALUE $SEX  M='?SKAWK?'
                  F='?DUWK?' ;
          VALUE SECC  1='?QKFRKSR&?'
                    2='?DB ?VKXM?'
                    3='?WKFYR&?'
                    4='?GIS? R&?';

      PROC SUMMARY;
      ④ CLASS SEX SEC AGE;
          ⑤ FORMAT SEX $SEX. SEC SECC.;
          ⑥ VAR HT WT;
          ⑦ OUTPUT OUT=MYDATA MEAN=MHT MWT SUM=SHT SWT;
      PROC PRINT DATA=MYDATA LABEL;
          LABEL SEX='?TJFQUF?'
                SEC='?VKXM?'
                AGE='?DUSFUD?'
                NAME='?DLFMA?'
                MHT='?VUDRBSTLSWKD?'
                MWT='?VUDRBSCPWND?'
                SHT='?TLSWKDGKQR&?'
                SWT='?CPWNDGKQR&?'
                _TYPE_='?QNSFB?'
                _FREQ_='?AUDTN?';

/*
//

```

- ① data 가 들어있는 LIBRARY의 member 를 명시해 준 JCL
- ② data step으로 SAS data set 은 DATA 1  
INFILE 은 ①에서 지정해준 INDATA로 이름을 명시하고  
INPUT문은 variable 을 나열해줌.
- ③ PROC FORMAT 문



- ④ PROC SUMMARY 문에서 CLASS 에 줄 data 들은 group data 이어야함.
- ⑤ 분석하고자 하는 변수들
- ⑥ OUTPUT OUT =MYDATA 이고 분석하고자 함은 평균(MEAN) 과 합(SUM) 이다.  
MEAN =뒤의 variable 들은 ⑤에서 명시한 variable 들의 평균 계산 값을 다시 변수로 주는 것임.  
SUM =뒤의 변수명도 같은 경우임.

\* 결과표는 table 3

< table 3 >

SAS										
13:46 FRIDAY, JULY 24, 1987										
OBS	시범 분류	파 분	역력 년	분 년	역사 영	평균 년	자 년	평균 년	시 년	계 년
1		.	.	0	26	162.846	56.7308	4234	1475	
2		.	25	1	1	160.000	50.0000	160	50	
3		.	26	1	2	158.500	51.0000	317	102	
4		.	27	1	1	160.000	54.0000	160	54	
5		.	28	1	1	168.000	54.0000	168	54	
6		.	29	1	2	156.000	52.5000	312	105	
7		.	30	1	3	160.000	55.3333	480	166	
8		.	31	1	4	166.500	58.2500	666	233	
9		.	32	1	3	163.000	60.0000	489	180	
10		.	33	1	3	166.000	58.0000	498	174	
11		.	38	1	2	166.500	59.0000	333	118	
12		.	41	1	2	163.500	59.5000	327	119	
13		.	45	1	1	157.000	56.0000	157	56	
14		.	48	1	1	167.000	64.0000	167	64	
15		바 가 계 분 년	.	2	7	164.571	58.8571	1152	412	
16		DB 파 분	.	2	8	160.375	53.2500	1283	426	
17		자 분 년	.	2	5	161.000	55.6000	805	278	
18		GIS 계	.	2	6	165.667	59.8333	994	359	
19		바 가 계 분 년	29	3	1	157.000	50.0000	157	50	
20		바 가 계 분 년	31	3	3	167.000	59.6667	501	179	
21		바 가 계 분 년	32	3	1	163.000	64.0000	163	64	
22		바 가 계 분 년	38	3	1	168.000	59.0000	168	59	
23		바 가 계 분 년	41	3	1	163.000	60.0000	163	60	
24		DB 파 분	25	3	1	160.000	50.0000	160	50	
25		DB 파 분	26	3	1	160.000	50.0000	160	50	
26		DB 파 분	28	3	1	168.000	54.0000	168	54	
27		DB 파 분	30	3	2	162.000	56.0000	324	112	
28		DB 파 분	33	3	1	150.000	45.0000	150	45	
29		DB 파 분	41	3	1	164.000	59.0000	164	59	
30		DB 파 분	45	3	1	157.000	56.0000	157	56	
31		자 분 년	26	3	1	157.000	52.0000	157	52	
32		자 분 년	27	3	1	160.000	54.0000	160	54	
33		자 분 년	30	3	1	156.000	54.0000	156	54	
34		자 분 년	31	3	1	165.000	54.0000	165	54	
35		자 분 년	48	3	1	167.000	64.0000	167	64	
36		GIS 계	29	3	1	155.000	55.0000	155	55	
37		GIS 계	32	3	2	163.000	58.0000	326	116	
38		GIS 계	33	3	2	174.000	64.5000	348	129	
39		GIS 계	38	3	1	165.000	59.0000	165	59	

< table 3 계속 >

40	여자	.	.	4	12	157.667	51.5000	1892	618
41	남자	.	.	4	14	167.286	61.2143	2342	857
42	여자	.	25	5	1	160.000	50.0000	160	50
43	여자	.	26	5	2	158.500	51.0000	317	102
44	여자	.	28	5	1	168.000	54.0000	168	54
45	여자	.	29	5	2	156.000	52.5000	312	105
46	여자	.	30	5	2	155.000	52.0000	310	104
47	여자	.	31	5	1	160.000	48.0000	160	48
48	여자	.	32	5	1	158.000	54.0000	158	54
49	여자	.	33	5	1	150.000	45.0000	150	45
50	여자	.	45	5	1	157.000	56.0000	157	56
51	남자	.	27	5	1	160.000	54.0000	160	54
52	남자	.	30	5	1	170.000	62.0000	170	62
53	남자	.	31	5	3	168.667	61.6667	506	185
54	남자	.	32	5	2	165.500	63.0000	331	126
55	남자	.	33	5	2	174.000	64.5000	348	129
56	남자	.	38	5	2	166.500	59.0000	333	118
57	남자	.	41	5	2	163.500	59.5000	327	119
58	남자	.	48	5	1	167.000	64.0000	167	64
59	여자	भाग	.	6	2	158.500	49.0000	317	98
60	여자	DB 파트	.	6	6	158.167	50.8333	949	305
61	여자	자	.	6	2	156.500	53.0000	313	106
62	여자	GIS	.	6	2	156.500	54.5000	313	109
63	남자	भाग	.	6	5	167.000	62.8000	835	314
64	남자	DB 파트	.	6	2	167.000	60.5000	334	121
65	남자	자	.	6	3	164.000	57.3333	492	172
66	남자	GIS	.	6	4	170.250	62.5000	681	250
67	여자	भाग	29	7	1	157.000	50.0000	157	50
68	여자	भाग	31	7	1	160.000	48.0000	160	48
69	여자	DB 파트	25	7	1	160.000	50.0000	160	50

< table 3 계속 >

1		SAS					13:46 FRIDAY, JULY 24, 1987			
OBS	서비스 종류	파라 미터	여객 년용	분 류	여수 일수	평균시차 일수	평균재고 일수	시작하계 일수	종료하계 일수	
70	여자	DB	파라	26	7	1	160.000	50.0000	160	50
71	여자	DB	파라	28	7	1	168.000	54.0000	168	54
72	여자	DB	파라	30	7	1	154.000	50.0000	154	50
73	여자	DB	파라	33	7	1	150.000	45.0000	150	45
74	여자	DB	파라	45	7	1	157.000	56.0000	157	56
75	여자		자라계	26	7	1	157.000	52.0000	157	52
76	여자		자라계	30	7	1	156.000	54.0000	156	54
77	여자	GIS	계	29	7	1	155.000	55.0000	155	55
78	여자	GIS	계	32	7	1	158.000	54.0000	158	54
79	남자		바가계	31	7	2	170.500	65.5000	341	131
80	남자		바가계	32	7	1	163.000	64.0000	163	64
81	남자		바가계	38	7	1	168.000	59.0000	168	59
82	남자		바가계	41	7	1	163.000	60.0000	163	60
83	남자	DB	파라	30	7	1	170.000	62.0000	170	62
84	남자	DB	파라	41	7	1	164.000	59.0000	164	59
85	남자		자라계	27	7	1	160.000	54.0000	160	54
86	남자		자라계	31	7	1	165.000	54.0000	165	54
87	남자		자라계	48	7	1	167.000	64.0000	167	64
88	남자	GIS	계	32	7	1	168.000	62.0000	168	62
89	남자	GIS	계	33	7	2	174.000	64.5000	348	129
90	남자	GIS	계	38	7	1	165.000	59.0000	165	59

4) 분산을 구하는 경우

```
//B19DMSU1 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//INDATA   DD DSN=E19.PROGRAM.LIBRARY(D6DATA),DISP=SHR,UNIT=DISK
//SYSIN    DD *
DATA DATA1;
  INFILE INDATA;
  INPUT  NAME $      SEX $      SEC AGE HT WT;
  PROC FORMAT;
    VALUE SECC 1='?QKFRKSR&?'
              2='DB ?VKXM?'
              3='?WKFYR&?'
              4='GIS? R&?';

  PROC SUMMARY;
    CLASS SEC ;
    FORMAT SEC SECC.;
    * VAR HT WT;
    ① OUTPUT OUT=MYDATA VAR=;
  PROC PRINT DATA=MYDATA LABEL;
    VAR SEC HT WT ;
    LABEL SEC='?VKXM?'
           HT='?QNSTKS(TLSWKD)?'
           WT='?QNSTKS(CPWND)?';
```

① 분산을 구하는 VAR을 사용하고 VAR=;이므로 새로운 변수명을 주지않고 \*에서 준 변수 HT WT를 같은 이름으로 계산결과 변수명으로 주겠다는 뜻임.

\* 결과표는 table 4

< table 4 >

OBS	파트 -	SAS	
		부사 (시자) LN LN 0	부사 (자점) LN 0
1	.	42.3754	44.0446
2	바가계 LN	31.6190	82.1429
3	DB 파트 -	46.2679	31.0714
4	자라계 LN	23.5000	22.8000
5	GIS 계	63.4667	21.3667

5) 변수를 선택해서 통계결과를 구하는 예

```

//B19DMSU2 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//INDATA   DD DSN=B19.PROGRAM.LIBRARY(D&DATA),DISP=SHR,UNIT=DISK
//SYSIN    DD *
DATA DATA1;
  INFILE INDATA;
  INPUT  NAME $    SEX $    SEC  AGE  HT  WT;
  PROC FORMAT;
    VALUE $SEX M='?SKAWK?'
           F='?DUWK?';
  PRCC SORT;
    BY SEX;
  PROC SUMMARY;
  ① BY SEX ;
  ② VAR HT WT;
  ③ OUTPUT OUT=MYDATA VAR(HT)=VHT;
  PROC PRINT DATA=MYDATA LABEL;
  ④ VAR SEX VHT ;
    FORMAT SEX $SEX.;
    LABEL SEX='?TJDQUF?'
           VHT='?QNSTKS(TLSWKD)?';
/*
//

```

- ① CLASS 문 사용대신 BY 문을 사용하여 group 별 분석 data 를 구하고자 함. BY 문을 쓰기 전에 반드시 PROC SORT 가 선행되어야 함.
- ② VAR 뒤의 HT WT 는 분석하고자하는 변수
- ③ VAR(HT)=VHT 로서 분산을 구하는데 ②에서 명시한 것 중 괄호속에 들은 것만 분산을 구하고 새로운 변수명 VHT 를 주었음.
- ④ PRINT 문에서 두 변수 SEX 와 VHT 만 OUTPUT 을 얻고자 함.

\* 결과표는 table 5

< table 5 >

SAS		
OBS	성 별	부사 (시 자) 신 장
1	여 자	18.7879
2	남 자	19.6044

6) 가중치를 주어서 통계결과를 구하는 예

```
//B19DMSU3 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC HANSAS
//SYSIN    DD *
DATA DATA1;
  INPUT    SEX $    WT NO;
  CARDS;
  M      45      4
  M      34      6
  F      56      5
  F      34      13
  M      24      10
  ;
PROC FORMAT;
  VALUE $SEX M='?SKAWK?'
          F='?DUWK?';
PROC SORT DATA=DATA1;
  BY SEX;
PROC SUMMARY DATA=DATA1;
  ① BY SEX;
  ② WEIGHT NO;
  ③ VAR WT;
  ④ OUTPUT OUT=MYDATA MEAN=;
PROC PRINT DATA=MYDATA LABEL;
  VAR SEX _FREQ_ WT;
  FORMAT SEX $SEX.;
  LABEL _FREQ_ ='?QLSEHTN?'
        SEX='?TJDQUF?'
```

① 그룹별로 결과를 원함.

BY문 사용 이전에 반듯이 PROC SORT가 선행 됨.

② WEIGHT 문 사용에 변수NO는 가중치로서 역할을 함.

③ 변수 WT를 사용하기 위해서

④ 평균을 구하고자 MEAN= ;을 주고 계산결과 변수명은 ③에  
서 선언한 WT를 그대로 사용

\* 결과표는 table 6



<table 6 >

OBS	서 비 예 리	SAS 비 다 스 트	평 균 치 수 준
1	여 자	2	40.1111
2	남 자	3	31.2000

## 마. PROC MEANS

1. 개 요 .....	109
2. Specifications .....	109
3. Option 들 .....	109
4. 실제 예제들 .....	111
1) output 의 소숫점 자리수 조정과 나누어 주는 관측치 조정 방법	111
2) 자동처리 통계 결과표 없이 User data set 으로 받아 처리하는 방 법.....	114
3) 예제에서 사용한 data .....	115

## 1. 개 요

PROC SUMMARY와 거의 유사하게 기술통계 결과를 얻을 수 있는데, output을 자동적으로 얻을 수도 있고 OUT SAS data set을 만들 수 있어 보고서 양식에 알맞은 PROC 이라서 예제를 들어 보았다.

## 2. Specifications

PROC MEANS Options ;

VAR variables ;

BY variables ;

FREQ variable ;

WEIGHT variable ;

ID variables ;

OUTPUT options ;

## 3. Option 들

1) PROC MEANS Options ;

여기서 option 가능한 것은

① DATA = SAS data set

② NOPRINT ..... 자동처리 통계결과표를 원하지 않을 때

③ MAXDEC = n ..... 결과표의 숫자 format을 정해줌

- ④  $VARDEF = DF \dots \dots$  분포가  $N - 1$   
 $= WGT \dots \dots$  분포가 가중치를 줌  
 $= N \dots \dots$  관측치가  $N$

2) OUTPUT Options

여기서 Options 가능한 것은

- ① OUT = SAS data set
- ② N  $\dots \dots$  관측치 수
- ③ NMISS  $\dots \dots$  결손치 갯수
- ④ MEAN  $\dots \dots$  평균
- ⑤ STD  $\dots \dots$  표준편차
- ⑥ MIN  $\dots \dots$  최소값
- ⑦ MAX  $\dots \dots$  최대값
- ⑧ RANGE  $\dots \dots$  범위
- ⑨ SUM  $\dots \dots$  합계
- ⑩ VAR  $\dots \dots$  분산
- ⑪ USS  $\dots \dots$  수정하지 않은 제품의 합
- ⑫ CSS  $\dots \dots$  수정한 제품의 합
- ⑬ CV
- ⑭ STDERR
- ⑮ T
- ⑯ PRT
- ⑰ SUMWGT 등이 가능함

#### 4. 실제 예제들

1) output 의 소숫점자리수 조정과 나누어 주는 관측치 조정방법

```

//B19DMME2 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN EXEC HANSAS
① //INDATA DD DSN=B19.PROGRAM.LIBRARY(D6DATA),DISP=SHR,UNIT=DISK
//SYSIN DD *
  ② DATA DATA1;
    INFILE INDATA;
    INPUT  NAME $    SEX $    SEC AGE HT WT;
  ③ PROC FORMAT;
        VALUE $SEX M='?SKAWK?'
              F='?DUWK?';
  ④ PROC SORT;
        BY SEX;
  ⑤ PROC MEANS MAXDEC=2 VARDEF=N;
  ⑥ BY SEX;
  ⑦ VAR HT WT;
  ⑧ OUTPUT OUT=MYDATA MEAN=MHT MWT;
  ⑨ PROC PRINT DATA=MYDATA LABEL;
  ⑩ FORMAT SEX $SEX.;
  ⑪ LABEL SEX='?TJDQUF?'
        MHT='?VUDRBS(TLSWKD)?'
        MWT='?VUDRBS(CPWND)?';
/*
//
  
```

- ① Data 가 들어있는 Data set 의 이름과 file 이름 ( INDATA )
- ② Data Step INFILE 뒤의 INDATA는 ①에서 준 file 이름
- ③ PROC FORMAT으로 변수 SEX를 남, 여로 바꾸고자 함.
- ④ SORT
- ⑤ PROC MEANS 에서 MAXDEC = 2 는 결과표를 소숫점 2 자리만 구하겠다고 줌, VARDEF = N 은 제수(divisor)를 관측치의 갯수만큼 하겠다는 option 임.
- ⑥ BY SEX도 group 별 통계 결과치를 얻을 목적으로 사용했는데 BY SEX 이전 반드시 ④에서 SORT가 이루어져야함..

- ⑦ 분석하고자 하는 변수들 ( HT, WT )
- ⑧ OUT = MYDATA 로 SAS data set 을 반드시 명시하고 구하고자 하는 기술통계 명령 을 주어야함.  
위의 예는 MEAN으로 평균을 구하고자 함.
- ⑨ OUTPUT 에서 형성한 SAS data set 인 MYDATA 를 Print 하고자 함.
- ⑩ FORMAT 문으로 ③에서 바꾸어준 것을 실제로 적용하는 문장
- ⑪ LABEL 로 변수를 한글로 바꾸어 줌.

\* 결과표는 table 1 ..... PROC MEANS 의 자동 output 결과  
table 2 ..... MYDATA 의 output 결과

< table 1 >

		SAS								
		16:13 FRIDAY, JULY 24, 1987								
VARIABLE	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE	STD ERROR OF MEAN	SUM	VARIANCE	C.V	
----- SEX=F -----										
HT	12	157.67	4.15	150.00	168.00	1.20	1892.00	17.22	2.6	
WT	12	51.50	3.10	45.00	56.00	0.89	618.00	9.58	6.0	
----- SEX=M -----										
HT	14	167.29	4.27	160.00	174.00	1.14	2342.00	18.20	2.5	
WT	14	61.21	5.18	54.00	75.00	1.39	857.00	26.88	8.4	

< table 2 >

SAS			
OBS	성별 이름	평균(식자) 오    나    노    오	평균(체중) 오    나    오
1	여자	157.667	51.5000
2	남자 오	167.286	61.2143

2) 자동처리통계 결과표없이 User data set 으로 받아 처리하는 방법

```

//B19DMME1 JOB CLASS=A,MSGCLASS=X,MSGLEVEL=(0,0)
//RUN      EXEC  HANSAS
//INDATA   DD DSN=B19.PROGRAM.LIBRARY(D6DATA),DISP=SHR,UNIT=DISK
//SYSIN    DD *
DATA DATA1;
  INFILE INDATA;
  INPUT  NAME $      SEX $      SEC  AGE  HT  WT;
  PROC FORMAT;
    VALUE SECC 1='?QKFRKSR&?'
            2='DB ?VKXM?'
            3='?WKFYR&?'
            4='?GIS? R&?';
  PROC SORT;
    BY SEC;
  ① PROC MEANS NOPRINT;
    BY SEC ;
  ② VAR HT WT;
  ③ OUTPUT OUT=MYDATA VAR=VHT VWT;
  ④ PROC PRINT DATA=MYDATA LABEL;
    FORMAT SEC SECC.;
    LABEL SEC='?VKXM?'
           VHT='?QNSTKS(TLSWKD)?'
           VWT='?QNSTKS(CPWND)?';
/*
//

```

- ① NOPRINT option으로 자동 output을 원하지 않음.
- ② 분석변수 HT WT를 줌.
- ③ OUTPUT OUT=MYDATA를 반드시 주고, 분산 VAR을 구 하겠다는 뜻임.
- ④ PROC PRINT로 SAS data set는 MYDATA임.

\* 결과표는 table 3



<table 3>

OBS	파트 -	SAS	
		부사 (시자) 년    년    0	부사 (재중) 년    0
1	바가계 년	31.6190	82.1429
2	DB    파트 -	46.2679	31.0714
3	자르계 년	23.5000	22.8000
4	GIS    계	63.4667	21.3667

3) 예제에서 사용한 data

THDTNSRHKS	M	1	32	163	64
RLATJFGML	F	1	31	160	48
QKRALGP	F	2	28	168	54
GHKDDUDTJR	M	3	27	160	54
TLSGHWND	M	3	31	165	54
DLEOGUD	M	4	33	174	64
AKRUDGML	F	4	32	158	54
RLAWLSQHR	F	3	26	157	52
GHDGPRUD	F	2	30	154	50
QKRGHDUS	F	2	33	150	45
FBWPWJD	F	1	29	157	50
DLAUDGH	M	1	31	167	56
DLDYDTJR	M	4	32	168	62
DHTKARE	M	2	30	170	62
QKDDBSGHK	M	1	38	168	59
GKSTKDRNJS	M	4	33	174	65
DLWNSGUD	M	1	41	163	60
RLAXOTJR	M	2	41	164	59
RLAGUDRNJS	M	3	48	167	64
GJSKARJ	M	4	38	165	59
TJDDMLTNS	F	2	45	157	56
DLRHKSDUD	M	1	31	174	75
QKRAUDGML	F	3	30	156	54
RLATJSSU	F	4	29	155	55
DKSRUDWK	F	2	25	160	50
RLAGMLDUD	F	2	26	160	50

PROC MEANS 에서 사용한 data로서 B19-PROGRAM. LIB-RARY 의 D6DATA 의 member 로 구성 되어 있음.

바. 한글자모 영문 code

한글자모	로마자	한글자모	로마자
ㄱ	R	ㅅ	\
ㄴ	S	ㅆ	-
ㄷ	E	ㅈ	K
ㄹ	F	ㅊ	I
ㅁ	A	ㅋ	J
ㅂ	Q	ㆁ	U
ㅅ	T	ㄱ	H
ㅇ	D	ㅋ	Y
ㅈ	W	ㅌ	N
ㅊ	C	ㅍ	B
ㅋ	Z	ㅑ	M
ㅌ	X	ㅓ	L
ㅍ	V	ㅕ	O
ㅎ	G	ㅗ	I
ㄱㄱ	Ø	ㅛ	P
ㄷㄷ	*	ㅜ	&
ㅂㅂ	@	ㅠ	\$