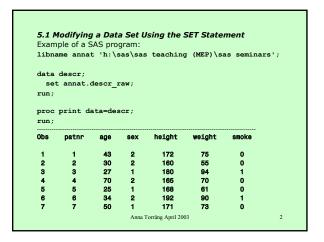
SAS seminar 2003-04-22 Anna Torrång, MEP Chpt 5: Modifying and Combining SAS Data Sets The Little SAS Book, 2nd ed., by Lora Delwiche & Susan Slaughter, SAS Publishing

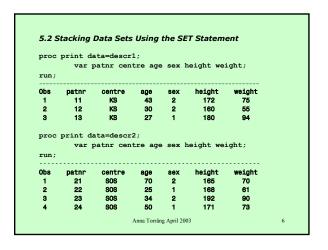
Anna Torrång April 2003

1

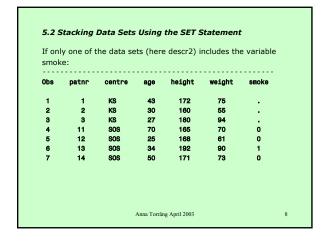


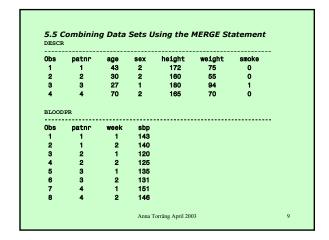
```
5.1 Modifying a Data Set Using the SET Statement
data descr;
 set annat.descr raw;
  if age<35 and sex=2;
  height_m=height/100;
  bmi=weight/(height_m**2);
  bmi=round(bmi,.1);
proc print data=descr;
  var patnr height weight bmi;
run;
      patnr
              height weight
                                 bmi
0bs
                160
192
                         55
90
                                 24.4
                      Anna Torrång April 2003
```

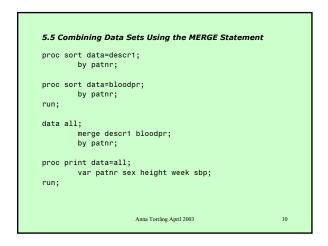
```
5.1 Modifying a Data Set Using the SET Statement
data descr cl:
  set descr;
    if bmi <= 24.9 then bmiclass=1;</pre>
         else if (25.0 \le bmi \le 29.9) then bmiclass=2; else if (30.0 \le bmi \le 34.9) then bmiclass=3;
         else if bmi => 35.0 then bmiclass=4;
run;
proc print data=descr_cl;
  var patnr height weight bmi bmiclass;
run:
                                    bmi bmiclass
0bs
                height weight
       patnr
                   160
192
                                     21.5
24.4
                             90
                         Anna Torrång April 2003
```



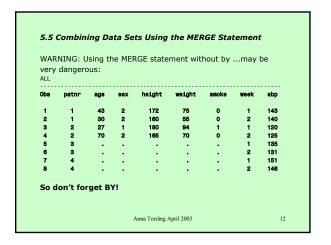
5.2 Stacking Data Sets Using the SET Statement set descr1 descr2; run; proc print data=all; var patnr centre age sex height weight; run; 0bs centre height age KS KS KS SOS SOS SOS 43 30 27 70 25 34 172 160 180 165 168 192 75 55 94 70 61 90 73 2 11 12 13 SOS 50 Anna Torrång April 2003







```
5.5 Combining Data Sets Using the MERGE Statement
Cont...
0bs
       patnr
               sex
                       height
                                  week
                                           sbp
                                           143
                                           140
120
                                    2
1
2
                          172
                          160
                          160
                                           125
                          180
180
                                    1
                                           135
                                           131
                          165
                                           151
                          165
                         Anna Torrång April 2003
                                                                 11
```



5.9 Using SAS Data Set Options

The most frequently used data set options:

Options: Tells SAS: KEEP = variable list variables to keep DROP = variable list variables to drop

RENAME = (oldvar = newvar) rename variable

FIRSTOBS = nstart reading at obs n

OBS = nstop reading at obs n

temporary variable for tracking whether that data set contributed to the current observation. IN = newvar-name

13

Anna Torrång April 2003

5.9 Using SAS Data Set Options data in1; set annat.descr_raw (keep = patnr age sex smoke); data in2; set annat.descr_raw (drop = height weight); data in3: set annat.descr_raw (rename = (height = heightcm)); proc print data=in3 (firstobs = 2 obs = 4); run: Anna Torrång April 2003 14

5.9 Using SAS Data Set Options

Tracking observations using the IN= option (IN creates a new temporary variable)

data all;

merge descr1 (in=indesc) bloodpr (in=bp); by patnr;

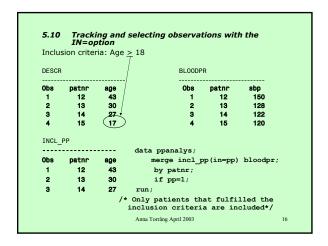
These variables exist only for the duration of current data step and are not added to the data set being created.

SAS gives IN=variables a value of 0 if that data set did not contribute to the current observation and a value of 1 if it

Anna Torrång April 2003

15

17



Tracking and selecting observations with the 5.10

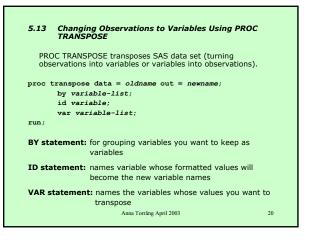
PPANALYS

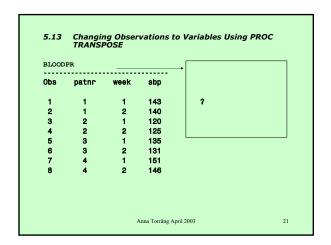
0bs	patnr	age	sbp
1	12	43	150
2	13	30	128
9	14	27	199

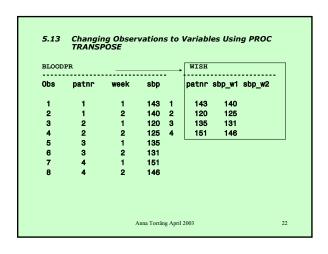
Anna Torrång April 2003

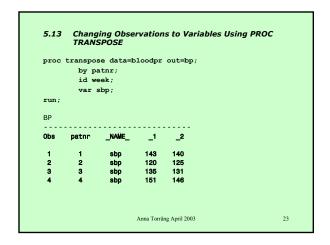
```
Writing Multiple Data Sets Using the OUTPUT
       Statement
RESULTS
       stud_nr
                  test
                           scores
                            56
40
89
60
         100
100
         100
101
         101
                             55
data score1 score2 score3;
         set results;
         if test=1 then output score1;
         if test=2 then output score2;
         if test=3 then output score3;
run;
SCORE1
       stud nr
                  test
                          scores
         100
101
                        56
60
Anna Torrång April 2003
                                                                18
```

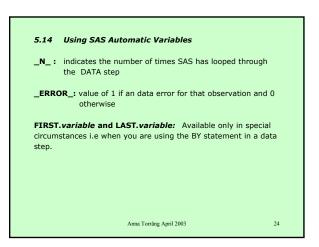
Usually SAS writes an observations from One Using the OUTPUT Statement Usually SAS writes an observation to a data set at the end of the DATA step, but you can override this default using the OUTPUT statement. This statement gives you control over when an observation is written to a SAS data set. data generate; DO loop with 6 iterations. do x=1 to 6; Without the OUTPUT, SAS would y=x***2; have written only 1 obs. output; at the end of the data step. end; run; Obs x y 1 1 1 1 2 2 4 3 3 9 4 4 16 5 5 25 6 6 36 Anna Torráng April 2003











```
5.14 Using SAS Automatic Variables

Example: FIRST.variable
CANCER

Obs lopnr diadat icd7

1 4567 19870508 170
2 6853 19840115 190
3 6853 19681108 170

proc sort data=cancer;
by lopnr diadat;

/* Include only the first diagnosis of cancer*/
data firstcan;
set cancer;
by lopnr diadat;
if first.lopnr;
run;

Anna Torráng April 2003 25
```

