SERIES #2   **date: 17th Feb 2006.**

Format Informat logic:

ABSTRACT:
         On applying the informat and format there are certain changes taking place in the dataset. With informat the changes are permanent while format is just an apparent change in the data of the dataset. Once we figure out this change in the dataset we understand the nature of informat or format.

What exactly is happening on applying Informat and Format ?

         We consider the following example to understand what happens on having an informat and format applied to a dataset one.

```sas
proc format;
      invalue $sex
            1='male'
            0='female';
      value $sex
            'male'=1
            'female'=0;
run;
/* generating the data set to work with the above informat */
data one;
      input sex $sex.;
datalines;
1
0
;
run;
```

         Here the sex variable is absorbed using the $sex informat hence the data sex would physically be 'male' and 'female' and not '1' and '0'. i.e on applying the informat the data is replaced (physically) by the informat data.

Now let us see the format code:

```sas
data one;
      set one;
      format sex $sex.;
run;
```

         Now the format makes the sex variable to contain '1' and '0' in place of 'male' and 'female'. But the format actually gives an *apparent* look to the data. Physically the data is stored as 'male' and 'female' and not as '1' and '0' though what we see in the data set is the (apparent) data '1' and '0'.

Now we prove the claim that format actually gives an apparent look to the data.
         We create a new variable newvar using the sex variable (since it seems that sex contains 1 and 0). If sex contains '1' and '0' then the newvar variable will contain '1' and '1' respectively. But if it doesn't contain the formatted data '1' and '0' but contains

'male' and 'female' then it would give an error, which is what happens. See the log error generated.

CODE:
```
data one;
      set one;
      value + sex;
run;
```

LOG:
```
194  data one;
195      set one;
196      value + sex;
197  run;
NOTE: Character values have been converted to numeric values at the places given by:
      (Line):(Column).
      196:13
NOTE: Invalid numeric data, sex='male' , at line 196 column 13.
sex=1 value=0 _ERROR_=1 _N_=1
NOTE: Invalid numeric data, sex='female' , at line 196 column 13.
sex=0 value=0 _ERROR_=1 _N_=2
NOTE: There were 2 observations read from the data set WORK.ONE.
NOTE: The data set WORK.ONE has 2 observations and 2 variables.
```

Informat conclusion:
> The informat takes the data ('1' & '0') from the raw data file or dataset or datalines and converts and stores it ('male' & 'female') physically in the new form permanently.

Format conclusion :
> The operation done on the data (sex) in the data step utilizes the actual data ('male' & 'female') in the variable and not the data ('1' & '0') which is the presented by the user or inbuilt format (here again we stress its an apparent view).

Now we investigate the various combinations of the informats and the formats that can be applied to the datasets data.

CODE for the investigation.
INFORMAT:

```
proc format;
      invalue $frmta 1='male' 0='female'; /*numeric --> character */
      invalue frmtb 1=11 0=10;/* numeric --> numeric */
      invalue $frmtc '1'='male' '0'='female';/* character --> character
: same as informat $frmta      */
      invalue frmtd 'male'=1 'female'=0; /* character--> numeric */
run;

/* format = $6.  informat $6. length  6 */
data one;
      input sex $frmta.;              /* numeric --> character */
datalines;
1
```

```sas
0
;
run;

/* format = best12.   informat 12. length   8 */
data two;
      input sex frmtb.;                /* numeric --> numeric */
datalines;
1
0
;
run;

/* this dataset three creation is same as the dataset one */
data three;
      input sex $frmtc.;               * character --> character*/
datalines;
1
0
;
run;

/* format = best12.   informat 12. length   8 */
data four;
      input sex frmtd.;                /* character--> numeric */
datalines;
male
female
;
run;
```

Conclusion : The informat (whether character or numeric informat) depends on the RHS data since the old data in the raw dataset or the datalines is being permanently replaced by the RHS value of the invalue statement in the proc format.

FORMAT CODE:

```sas
proc format ;
      value frmtw 1='male'  0='female';/* numeric --> character*/
      value frmtx 1=11 0=10;/* numeric --> numeric*/
      value $frmty 'male'=1 'female'=0;/* character--> numeric */
      value $frmtz 'male'='men' 'female'='women'; /* character -->
character */
run;

/* format frmtw6. informat 6 length 8*/
data one1;
      input sex ;
      format sex frmtw.;
datalines;
1
0
```

```
;
run;

/* format frmtx2. informat 2 length 8*/
data two1;
      input sex;
      format sex frmtx.;
datalines;
1
0
;
run;

/* format $frmty1. informat 1 length 8*/
data three1;
      input sex $;
      format sex $frmty.;
datalines;
male
female
;
run;

/* format frmtx5. informat 5 length 8*/
data four1;
      input sex $;
      format sex $frmtz.;
datalines;
male
female
;
run;
```

Conclusion : The format (whether character or numeric format) depends on the LHS data (in the value statement), since the old data in the raw dataset or the datalines is apparently being replaced by the LHS value of the value statement in the proc format.

```
Sachin Gadkar.
Cognizant Tech.Soln.
```