

STAT 541

Chapter 13
supplement:
Review of
Formatting Data

Informats

(**input**-related)

how to **read**
data values

CONVERT

[**interpretation**]

Formats

(**output**-related)

how to **write**
data values

PRINT

[**appearance**]

are instructions

Standard

examples:

2.1

\$5.

\$char5.

yymmdd8.

User-

Defined



PROC FORMAT

Formats and Informats

Standard Informat for Input

original values: 00 10 15 20 25 30 35 40
↓
desired values: 0.0 1.0 1.5 2.0 2.5 3.0 3.5 4.0

```
data grades;  
input grade;  
grade=grade/10;  
cards;  
20  
;
```

```
data grades;  
input grade 2.1;  
cards;  
20  
;
```

Standard Informat for Input

```
data amount;  
input amount1 dollar9.2 @1 amount2 9. @1 amount3 $9.;  
cards;  
$5,251.12  
$251.12  
251.12  
251.345  
;  
proc print; sum amount1-amount2;
```

Obs	amount1	amount2	amount3
1	5251.12	.	\$5,251.12
2	251.12	.	\$251.12
3	251.12	251.120	251.12
4	251.35	251.345	251.345
	=====	=====	
	6004.71	502.465	

Standard Informat for Fixed-Width Input

```
data one;
input title $19. author $18. stars 2. (rating1-rating3) ($1.+1) ;
cards;
Gone with the Wind Margaret Mitchell 5 A A B
Roots Alex Haley 5 A A B
Joy Luck Club Amy Tan 5 A A B
;
```

Obs	title	author	stars	rating1	rating2	rating3
1	Gone with the Wind	Margaret Mitchell	5	A	A	B
2	Roots	Alex Haley	5	A	A	B
3	Joy Luck Club	Amy Tan	5	A	A	B

```
input @20 author $18. @44 rating3 $1.;
```

Obs	author	rating3
1	Margaret Mitchell	B
2	Alex Haley	B
3	Amy Tan	B

Standard Informat for Delimited Input

```
data one;
infile cards delimiter=',';
input title :$19. author :$18. stars :2. (rating1-rating3) (:$1.) ;
cards;
```

The : argument indicates length is up to 19 characters.

```
Gone with the Wind,Margaret Mitchell,5,A,A,B
Roots,Alex Haley,5,A,A,B
Joy Luck Club,Amy Tan,5,A,A,B
;
```

with the LENGTH statement

```
length title $19. author $18. ;
infile cards delimiter=',';
input title author stars (rating1-rating3) ($);
```

The SAS System

Obs	title	author	stars	rating1	rating2	rating3
1	Gone with the Wind	Margaret Mitchell	5	A	A	B
2	Roots	Alex Haley	5	A	A	B
3	Joy Luck Club	Amy Tan	5	A	A	B

Naming Conventions for User-Defined Formats (VALUE, PICTURE) and Informats (INVALUE) in PROC FORMAT

Use valid SAS names that **do not end in a number.**

Character

Begins with \$

INFORMAT

+ up to 30 characters

FORMAT

+ up to 31 characters

Numeric

INFORMAT

Up to 31 characters

FORMAT

Up to 32 characters

VALID

\$gender

\$gender2F

INVALID

\$gender2

\$2gender

VALID

gender

gender2F

INVALID

gender2

2gender

In SAS code, refer to them with a period following their name,
BUT do not use the period in PROC FORMAT.

User-Defined Informat for Input

```
proc format;  
  invalue $gender 1='F' 2='M';  
  
data one;  
  input gender $gender.;  
  cards;  
  1  
  2  
  ;
```

The SAS System

Obs	gender
1	F
2	M

Data Validation with Informats

```
proc format;  
  invalue $check M,F=_same_  
             other=_error_ ;  
data one;  
input gender $check.;  
cards;  
M  
F  
Z  
;
```

The SAS System

Obs	gender
1	M
2	F
3	

Standard and User-Defined Formats for Output

Appearance of Numbers in Output

```
data grades;
input ssn;
cards;
056789012
;

proc format;
picture s low-high=9999999999
      (prefix='SSN: ');
proc print;
  var ssn;  format ssn s.;
proc print;
  var ssn;  format ssn ssn.;
```

The SAS System

Obs	ssn
1	SSN: 056789012

The SAS System

Obs	ssn
1	056-78-9012

**Why are formats and
informats useful for SAS
date variables?**

Date type ⇒ **an integer equal to the number of days elapsed since Jan. 1, 1960**

Date	SAS Date Value
Dec. 31, 1959	-1
Jan. 1, 1960	0
Jan. 2, 1960	1
Jan. 10, 1960	9

Appearance of SAS Dates (Numbers) in Output

```
data one;
input dob mmddyy.;
yymmdd8=dob;
mmddyy10=dob;
cards;
123 159
010160
0102 60
011060
;
proc print;
  format yymmdd8 yymmdd8.
         mmddyy10 mmddyy10.;
```

Obs	dob	yymmdd8	mmddyy10
1	-1	59-12-31	12/31/1959
2	0	60-01-01	01/01/1960
3	1	60-01-02	01/02/1960
4	9	60-01-10	01/10/1960

Appearance of Numbers in Output

```
data one;  
input a b c;  
a3dot0=a; b3dot1=b; c4dot2=c;  
cards;  
3.4 3.44 3.444  
3.5 3.45 3.445  
3.6 3.46 3.446  
;
```

```
proc print; var a a3dot0 b b3dot1 c c4dot2;  
format a3dot0 3.0 b3dot1 3.1 c4dot2 4.2;
```

Obs	a	a3dot0	b	b3dot1	c	c4dot2
1	3.4	3	3.44	3.4	3.444	3.44
2	3.5	4	3.45	3.5	3.445	3.45
3	3.6	4	3.46	3.5	3.446	3.45

Formats for Output: Just a PROC

```
proc format;  
  value gender 1=The Women 2=The Men;  
  
data one;  
  input gender;  
  cards;  
  1  
  2  
  ;  
  
proc freq;  
  tables gender/nocum nopct;  
  format gender gender.;
```

The SAS System

The FREQ Procedure

gender	Frequency
The Women	1
The Men	1

Formats for Output: Many PROCs

```
proc format;  
  value gender 1=The Women 2=The Men;  
data one;  
input gender;  
format gender gender.;  
cards;  
1  
2  
;  
proc print;  
proc freq;  
  tables gender/nocum nopct;
```

The SAS System

Obs	gender
1	The Women
2	The Men

The SAS System

The FREQ Procedure

gender	Frequency
The Women	1
The Men	1

Numeric Missing Values

A numeric missing value in SAS means there is no data value.

Missing Value Type	Representation	Description
Regular Numeric	.	Single period
Special Numeric	.a .b .c .d .e .f .g .h .i .j .k .l .m .n .o .p .q .r .s .t .u .v .w .x .y .z	Single period followed by a letter These are not case-sensitive. (.A is equivalent to .a).
Special Numeric	._	Single period followed by an underscore

Meaning of Special Numeric Missing Values

```
data one;
input x @@;
cards;
.i .a .
;

proc sort data=one;
  by x;

proc format;
  value sample 1=yes
              0=no
              .=no response
              .i=illegible
              .a=attrition;

proc sort data=one; by x;
proc print data=one;
  format x sample.;
```

The SAS System

Obs	x
1	no response
2	attrition
3	illegible

Group Processing

```
data one;
input percentile @@;
cards;
1 2 26 31 51 52 75 99
;

proc format;
  value group 1-25='Q1 (1-25)'
              26-50='Q2 (26-50)'
              51-75='Q3 (51-75)'
              76-99='Q4 (76-99)';

proc freq;
  tables percentile/nocum nopct;
  format percentile group.;
  label percentile='quartile';

run;
```

The SAS System

The FREQ Procedure

quartile	
<u>percentile</u>	<u>Frequency</u>
Q1 (1-25)	2
Q2 (26-50)	2
Q3 (51-75)	3
Q4 (76-99)	1

```

proc format;
  invalue number
    'A'=4 'B+'=3.5
    'B'=3 'C+'=2.5
    'C'=2 'D+'=1.5
    'D'=1 'F'=0;
  value $stu
    'A'='A student'
    'B+', 'B'='B student'
    'C+', 'C'='C student'
    'D+', 'D'='D student'
    'F'='F student';

```

```

data grades;
input lettergrade $;
numgrade = input(lettergrade, number.);
student = put(lettergrade, $stu.);
cards;
A
;

```

Create new variables from existing ones (recode) with PUT and INPUT functions

The SAS System			
Obs	lettergrade	numgrade	student
1	A	4	A student

PUT Function and Format

General syntax without optional arguments:

```
PUT (source, format)
```

Always returns a **character value** by applying a format to an expression (*source*)

Converts numeric to character values

The format must be of the same type as *source*.

INPUT Function and Informat

General syntax without optional arguments:

```
INPUT (source, informat)
```

Returns a value by applying an informat to an expression (`source`)

Informat type determines numeric or character type result.

Converts character to numeric values

Array Index Values are Easier to Follow

Points	Level				
Color	1	2	3	4	5
red	4	5	5	5	6
green	3	4	5	5	6
Blue	2	3	4	5	6

```
proc format;
  invalue num red=1 green=2 blue=3;
data one;
input color $ level;
array grid
  {1:3, 1:5} _temporary_
  (4 5 5 5 6
  3 4 5 5 6
  2 3 4 5 6);
points=grid(input(color,num.), level);
cards;
red 4
;
```

The SAS System

Obs	color	level	points
1	red	4	5

LOOKUP TABLES

```
data universe;
input ssn;
cards;
111111111
123456789
987654321
222222222
;
proc format;
value lookup
  123456789=keep
  987654321=keep;
data subset;
set universe;
if put(ssn, lookup.)='keep';
```

The SAS System

Obs	ssn
1	123456789
2	987654321

Q: What do I do if there's A LOT to type??

```
proc format;
  value names
    0000000001='Aguinaldo, Emilio'
    0000000002='Rizal, Jose'
    .
    .
    .
    0000099999='Marcos, Ferdi';
```

A: If the information is in a data set, you can create the format automatically.

INPUT CONTROL DATA SETS (CNTLIN=)

```
data one;
input ssn $9. name $;
cards;
123456789 Mickey
123456789 Mickey
;

data formatssn;
set one;
fmtname='$person';
type='C';
rename ssn=start name=label;

proc sort data=formatssn
  out=formatssn nodupkey;
  by start;

proc format cntlin=formatssn;
```

TYPE:

C for Character FORMAT

N for Numeric FORMAT

I for Numeric INFORMAT

J for Character INFORMAT

```
proc format;
  value $name
    '123456789'='Mickey';
proc print data=one;
  var ssn;
  format ssn $name.;
proc print data=one label;
  var ssn;
  format ssn $person.;
  label ssn='Celebrity';
```

The SAS System

Obs	Celebrity
1	Mickey
2	Mickey

OUTPUT CONTROL DATA SETS (CNTLOUT=)

```
proc format cntlout=outformat;  
  value $gender 'M'='male'  
              'F'='female';
```

The SAS System 16:18 Thur

Obs	FORMAT NAME	START	END	LABEL	MIN	MAX	DEFAULT	LENGTH	FUZZ	PREFIX	MULT	FILL	NOEDIT	TYPE	SEXCL	EXCL	HLO
1	GENDER	'	'	'female'	1	40	24	24	0		0		0	C	N	N	
2	GENDER	'M'	'M'	'male'	'F'	1	40	24	24	0		0	0	C	N	N	

The CONTENTS Procedure

Data Set Name	WORK.OUTFORMAT	Observations	2
Member Type	DATA	Variables	21
Engine	V9	Indexes	0
Created	Thursday, April 03, 2008 04:18:59 PM	Observation Length	128
Last Modified	Thursday, April 03, 2008 04:18:59 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	12288
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	95
Obs in First Data Page	2
Number of Data Set Repairs	0
File Name	C:\DOCUME~1\IGo\LOCALS~1\Temp\SAS Temporary Files_TD4424\outformat
Release Created	9.0101M3
Host Created	XP_PRO

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
20	DATATYPE	Char	8	Date/time/datetime?
18	DECSEP	Char	1	Decimal separator
7	DEFAULT	Num	3	Default length
19	DIG3SEP	Char	1	Three-digit separator
16	EEXCL	Char	1	End exclusion
3	END	Char	3	Ending value for format
12	FILL	Char	1	Fill character
1	FMTNAME	Char	32	Format name
9	FUZZ	Num	8	Fuzz value
17	HLO	Char	11	Additional information
4	LABEL	Char	24	Format value label
21	LANGUAGE	Char	8	Language for date strings
8	LENGTH	Num	3	Format length
6	MAX	Num	3	Maximum length
5	MIN	Num	3	Minimum length
11	MULT	Num	8	Multiplier
13	NOEDIT	Num	3	Is picture string noedit?
10	PREFIX	Char	2	Prefix characters
15	SEXCL	Char	1	Start exclusion
2	START	Char	3	Starting value for format
14	TYPE	Char	1	Type of format

PROC CONTENTS
labels describe
variables in output
control data sets.

User-defined formats can be stored in format catalogs and accessed later.

PC SAS Example

```
libname flib 'C:\Documents and Settings\IGo\Desktop\MARS';  
proc format library=flib;  
**numbering Wayne used = new numbering;  
value $fixgrp '04'='05B'  
              '05'='04W'  
              '06'='07H'  
              '07'='06AP'  
              '08'='08AI';
```

C:\Documents and Settings\IGo\Desktop\MARS

Name	Size	Type	Date Modified
lisrel		File Folder	4/9/2008 4:01 PM
references		File Folder	4/9/2008 4:23 PM
2006-07 PACT ELA...	496 KB	Microsoft ...	7/6/2007 9:09 AM
formats.sas7bcst	17 KB	SAS Catalog	4/11/2008 8:32 AM

SAS - [CATALOG]

File Edit View Tools Solutions Window Help

Command ==>
Contents of 'Tmp1.Formats'

Name	Size	Type	Modified
Fixgrp	0.3KB	Formatc	11Apr08:08:26:41

Later:

```
options fmtsearch = (flib);
```

NESTED FORMATS

```
proc format;
  value one
    3-3.5=good or very good
    2-2.5=average or above average
    1-1.5=poor or below average;
  value two
    4=excellent
    3-3.5=good or very good
    2-2.5=average or above average
    1-1.5=poor or below average
    0=failure;
  value nested 4=excellent
    1-3.5=(|one.|)
    0=failure;

data one;
input x1 x2 xnested;
cards;
3.5 3.5 3.5
4 4 4
;

proc print;
format x1 one. x2 two. xnested nested.;
```

The SAS System

Obs	x1	x2	xnested
1	good or very good	good or very good	good or very good
2		4 excellent	excellent