Handling missing values in you

READING MISSING DATA USING SAS USING A "FORMAT" STATEMENT

1. Define the column width of each variable.

```
data new1 ;
   input @1 year 3. @10 foodcons 3. @18 pretail 4. @26 dispinc 4.;
cards;

546               8456
         111                5555
;
run;
proc print;
run;
```

(How to read data above, assuming 36 columns, there can be more):

									(10)									(20)									(30)						
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
5	4	6															8	4	5	6															
									1	1	1														5	5	5	5							

The (3.) and (4.) in the **INPUT** statement refer to the column width of each variable in question: 3 columns and 4 columns, respectively. Note that if you had data with columns appearing behind a decimal point, such as a GPA of 3.46, then your column specification would be 4.2: 4 total columns, including the decimal point, with two of those four columns consisting of values behind (to the right of) the decimal point.

2. Delimit the data (separate data fields with a comma)

```
data new2;
 infile cards dsd;
 input year foodcons pretail dispinc;
 cards;
546,,8456,
,111,,5555
;
run;
proc print data=new2;
run;
```

DSD (delimiter-sensitive data)

DSD option sets the default delimiter to a comma.

SAS treats two consecutive delimiters as a missing.

Delimiter='list-of-delimiting-characters'

specifies one or more characters to read as delimiters.

One needs to specify the delimiter if it is not comma.

```
data new3;
 infile cards dsd delimiter='a';
 input year foodcons pretail dispinc;
 cards;
546aa8456a
a111aa5555
;
run;
proc print data=new3;
run;
```

RECODING VARIABLE VALUES INTO MISSING VALUES IN SAS

We can recode numeric zeros into SAS system missing by using an **IF-THEN** statement in SAS. SAS uses the period symbol ('.') as its missing value identifier.

The following SAS program convert numeric zero to SAS system missing values. We can do the same to convert 99, -99 or any other values to missing.

```
data recode;
 input id a b c;
 if a=0 then a=.;
 if b=0 then b=.;
cards;
1 2 3 4
2 0 4 5
3 4 0 6
;
proc print data=recode;
run;
```

MEAN SUBSTITUTION FOR MISSING VALUES IN SAS

SAS has a procedure called **PROC STANDARD** that can be used to standardize some or all of the variables in a SAS data set to a given mean and/or standard deviation and produce a new SAS data set that contains the standardized values. In addition, there is a **REPLACE** option that substitutes all missing values with the variable mean.

The following SAS code demonstrates the use of **PROC STANDARD** for mean substitution.

```
data raw ;
    input v1-v10 ;
cards;
1 1 1 1 1 . 1 1 1 1
2 2 2 . 2 . 2 2 2 2
3 3 3 3 3 3 . . 3 3
4 4 4 . . 4 4 4 4 4
5 5 5 5 5 5 5 5 . .
;
proc standard data=raw out=stnd replace print;
   var v1-v10;
proc print data=stnd;
run;
```