
34

Opening the Dataset: A Twelve-Step Program for Dataholics

Appendix 34.A Dataset IN

```
data IN;  
input ID $1. X1 X2 X3 X4;  
cards;  
1 1 2 3 4  
2 1 2 3 .  
3 1 2 . .  
4 1 . . .  
5 . . . .  
;  
run;
```

Appendix 34.B Samsize Plus

```
PROC CONTENTS data=IN noprint  
out=out1(keep=libname memname nobs type); run;
```

```
PROC FORMAT;
value typefmt 1='Numeric' 2='Character';
run;
```

```
PROC SUMMARY data=out1 nway;
class libname memname type;
id nob;
output out=out2
(drop=_type_ LIBNAME MEMNAME
  rename=(freq_=NUMBER_of_VARIABLES NOBS=SAMPLE_SIZE));
format type typefmt.;
run;
```

```
PROC PRINT data=out2 noobs;
run;
```

Appendix 34.C Copy-Pasteable

```
PROC CONTENTS data=IN
out = vars (keep = name type)
noprint;
run;
```

```
PROC SQL noprint;
select name into: varlist_is_here separated by ''
from vars;
quit;
%put _global_;
```

Appendix 34.D Missings

```
PROC SUMMARY data=IN;
var x1 x2 x3 x4;
output out=out3(drop=_type_ rename=(_freq_=sam_size)) nmiss=n_miss1-n_miss4;
run;

data out4;
set out3;
array nmiss n_miss1-n_miss4;
array pct_miss PCT_MISSING_X1-PCT_MISSING_X4;
do over nmiss;
pct_miss= nmiss/sam_size;
end;
keep PCT_MISSING_X1-PCT_MISSING_X4;
run;

PROC PRINT data=out4;
format PCT_MISSING_X1-PCT_MISSING_X4 PERCENT8.1;
run;
```