

# SAS Cheat Sheet for Stat-342

Carl James Schwarz

April 9, 2016

## 1 Reading data files into SAS dataset

```
DATA dsname1(dsoptions)
    dsname2(dsoptions) ....;
INFILE filelocation infileoptions;
LENGTH cvar1 $length1
        cvar2 $length2 ....;
INPUT var1 cvar2 var3 cvar4 ....;
ATTRIB var1 LABEL=    FORMAT=    ;
... processing statements ....
run;
```

### 1.1 Important INFILE options

- MISSOVER
- DLM= DSD
- FIRSTOBS=

### 1.2 Common INFORMATS

Use the : format modifier with these formats.

- COMMAw.d numeric values with commas
- \$ character data, but don't forget the LENGTH
- ANYDATEw accommodates a variety of dates but look at documentation!

### 1.3 Common out FORMATS

- w.d standard numeric
- \$n character
- COMMAw.d numeric values with commas
- DATEw.d,YYMMDDw. - use the ISO standard with 4 digit years!

### 1.4 Format modifiers

Used typically with list input to modify the format attached to a variable.

- (colon) : typically used for numeric variables for dates/times/commas etc.
- (ampersand) & used for character values with embedded blanks

## 2 Importing data from database systems

```
Proc IMPORT FILE=filename
            OUT=dsname DBMS=dbms REPLACE;
            GUESSINGROWS=nnnn;
            GETNAMES=yes;
run;
```

Common dbms are csv.

## 3 Modifying existing SAS datasets

### 3.1 Subsetting observations

```
DATA dsname;
    SET dsname;
    IF condition ;
    IF condition then DELETE;
    .... statements ....
run;
```

### 3.2 Selecting variables

```
DATA dsname;
    SET dsname;
    ...
    KEEP var1 var2 ...;
    DROP var1 var2 ....;
run;
```

### 3.3 Merging datasets

```
Proc SORT data=DS1; by bvar1 bvar2 ...; run;
Proc SORT data=DS2; by bvar1 bvar2 ...; run;
DATA both;
    MERGE ds1 ds2 ....;
    BY bvar1 bvar2 ....;
run;
```

What happens if records are in one dataset but not the other? Refer to manuals for use of IN= variables to keep track of which dataset is active in the merge.

### 3.4 Stacking datasets

```
Proc SORT data=DS1; by bvar1 bvar2 ...; run;
Proc SORT data=DS2; by bvar1 bvar2 ...; run;
DATA both;
    SET ds1 ds2 ....;
    BY bvar1 bvar2 ....;
run;
```

### 3.5 Derived variables and functions

SAS has an extensive list of function. See the help file for details.

Useful functions are:

- day(date), month(date), year(date), weekday(date), etc.
- index(text, string)
- max(var1, var2, ....), min(), sum(), mean() etc.
- round(var)

- substr(test, begin, length) - differs from C
- upcase(), lowcase() - change case of text
- word(string, n)

## 4 Graphical Procedures

```
Proc SGPlot data=dsname1;
  SCATTER X= Y= / GROUP=;
  HIGHLOW X= HIGH= LOW= ;
  XAXIS label=
    order=
    offset=(left, right);
```

Check manual for many other options.

## 5 Reporting Procedures

### 5.1 PRINT

```
Proc PRINT DATA=dsname (OBS=nnn)
  LABEL SPLIT NOOBS;
  VAR var1 var2 ...;
  ATTRIB var1 LABEL= FORMAT= ;
  PAGEBY var;
run;
```

### 5.2 TABULATE

```
Proc TABULATE DATA=dsname MISSING;
  CLASS pagevar1 rowvar2 rowvar3 colvar4 ...;
  VAR analvar ...;
  TABLE pagevar1,
    rowvar2*rowvar3,
    colvar4*analvar*(N*F=w.d MEAN*F=w.d...); BY bvar1 bvar2 ...;
run;
```

## 6 Analysis Procedures

### 6.1 FREQ

```
Proc Freq data= ;
  table v1 * v2 / chisq nocol nopercnt;
run;
```

### 6.2 GENMOD

```
Prog GENMOD data=...;
  class group;
  model y = group / dist=binomial
    link=logit type3;
  lsmeans group / cl diff
    adjust=tukey ilink;
  ods output lsmeans=.....;
```

See also LOGISTIC for logistic regression models;

### 6.3 GLIMMIX

### 6.4 GLM

Avoid using GLM for any linear mixed models. Use MIXED.

```
proc GLM data= ...;
  class ...
  model y = x;
  lsmeans x / lines cl pdiff adjust=tukey;
  ods output lsmeans=.....;
run;
```

### 6.5 MEANS

```
Proc SORT data=dsname;
  BY bvar1 bvar2 ...; run;
Proc MEANS DATA=dsname NOPRINT;
  VAR var1 var2 ...;
```

```
OUTPUT OUT=
  statistic(var)=name ....;
run;
```

Popular statistics are *n*, *mean*, *stddev*, *stderr*, *lclm*, *uclm*. See also SUMMARY.

### 6.6 MIXED

### 6.7 REG

```
Proc REG data=....1
  model y= x1 x2 x3 / clb;
run;
```

### 6.8 SURVEYSELECT

```
Proc surveyselect data=population
  out=sample
  method=
  sampsiz= samprate=
  seed=
  outhits
  reps=;
run;
```

Common methods are *srs* and *urs*. Specify the *sampsiz=* or *samprate=* but not both. Many other methods and ways to specify sampling (e.g. if clustering exists) can be specified. Many other similar procedures for analysis of survey data.

### 6.9 TTEST

```
Proc TTEST data=....1
  class group;
  var var1; /* independent sample */
run;
```

Always use the Welch version of the *t*-test. *Paired* statement used for paired t-test or use *Proc Univariate* on difference.

### 6.10 UNIVARIATE

```
Proc UNIVARIATE data=.... cibasic robustscale;
  var var1 var2 ....;
  output out= statistic(variable)= ;
run;
```

Generate lots of output! Can also generate histograms etc, but I prefer *SGplot*.

## 7 Split-Apply-Combine

The BY statement can be used with any procedure along with ODS tablename=dsname to send selected output to a SAS dataset.

```
Proc SORT data=dsname; by ....;
Proc BLAH data=dsname;
  BY bvar1 bvar2;
  ... statements...
  ODS tablename=newds;
run;
```

## 8 Wide-Long and Long-Wide

```
data long;
```

```
set wide;
length test $10.;
grade = t1; test="Test1"; output;
grade = t2; test="Test2"; output;
....
run;
```

```
proc transpose data=long out=wide;
  by studentnumber;
  var grade;
  id test;
run;
```

## 9 Sending SAS output to other destinations

Refer to manual for extensive help on using ODS. Common usage to create PDF or MSWord document.

```
ODS PDF FILE='filename.pdf' style=.....;
.... procedures that generate output ...
ODS PDF CLOSE;
ODS RTF FILE='filename.rtf' style=.....;
.... procedures that generate output ...
ODS RTF CLOSE;
```

## 10 Generating random numbers

SAS has a complete set of function to generate pseudo-random numbers. Some examples are:

- ranuni(seed); rand('uniform'); a U[0,1] with E=.5 and  $SD=\sqrt{1/12}$ .
- rannor(seed); rand('normal'); a N(0,1) with E=0, and SD=1;
- rand('lognormal') with  $E=e^{.5}$ , and  $SD=\sqrt{(e-1)e}$ .

## 11 Bootstrapping

General procedures for CRD/SRS and statistics related to the mean.

- Resample *K* times with replacement from the original dataset using *Proc SurveySelect* and the *method=urs*, *sampfrac=1*, *outhits* and *reps=* options.
- Compute estimate for each bootstrap sample.
- Look at bootstrap sampling distribution to compute relevant quantities of interest.

## 12 Macro Variables

%LET mvar= ...; /\* sets the macro variable \*/  
Use &mvar where you want to replace the macro variable by the assigned text. Careful of quotes, i.e. "&mvar" vs '&mvar'.