Friday, December 7, 2018

SAS Programming

Stacey Syphus, senior manager

- Week 1
  - SAS Studo web
  - Syntax:
    - data: manipulation
    - proc: perform
    - run: execute steps
    - statements end with ;
    - global: title, options, libname;
    - comments: /* … */ or * … ;
  - format code button!
- Week 2
  - accessing data: structure(rows and columns: sas, oracle, excel, hadoop; engine read), unstructure(no define columns: txt, delimited, json; import only)
  - column attributes: name, type(date: 01Jan1960), length
  - SAS Library: LIBNAME *libref engine* “path”;
    - engine: base, excel, teradata, hadoop, xlsx, …
    - path: directory, or DB connection if server based SAS
    - native: work(temp, auto, default), sashelp
    - libref clear; - disconnect and unlock library
  - XLSX: require SAS access to PC
    - options:
      - validvarname=V7(_ replace space, limit to 32 char),

```
FILENAME REFFILE '/home/ymtan10/EPG194/data/class.xlsx';

PROC IMPORT DATAFILE=REFFILE
    DBMS=XLSX
    OUT=WORK.IMPORT;
    GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.IMPORT; RUN;
```

- activity/p102a01.sas

```
libname t1 xlsx "s:/workshop/data/class.xlsx";
proc contents data=t1.class_birthdate; run;
```

- build-in import wizards: EG, SAS Studio, SAS Windowing environment
- proc import:
  - <REPLACE>
  - <GUESSINGROWS=n|MAX>: use for attibutes
  - xlsx reads first sheet, or use sheet= for others.

```
PROC IMPORT DATAFILE="path/filename" DBMS=filetype
            OUT=output-table;
RUN;
```

- Week 3: explore data
  - proc: print, means[num], univariate[num], freq, sort by VAR; sements report
  - where expressions;
    - subsetting [IN, NOT IN](,), AND/OR, =/EQ
    - date: “ddMMMyyyy”d;
    - between: 0<salary<1mil; between 0 and 1mil;
    - missing: VAR is missing; VAR is not missing; VAR is null;
    - var LIKE “value”; %-any number of characters; _-single character;
  - macro variables: %LET var=val; &var
  - formatting: <$>fname<w>.<d>
    - $ for character, w for width, d for digits after decimal
    - multiple: format var1 var2 f1. var3 f2;
  - sorting:
    - <NODUPKEY NODUPRECS <dupout=table1>>; BY _ALL_;

```
PROC PRINT DATA=input-table(OBS=n);
    VAR col-name(s);
RUN;

PROC MEANS DATA=input-table;
    VAR col-name(s);
RUN;

PROC UNIVARIATE DATA=input-table;
    VAR col-name(s);
RUN;

PROC FREQ DATA=input-table;
    TABLES col-name(s);
RUN;
```

```
PROC SORT DATA=input-table <OUT=output-table>;
    BY <DESCENDING> col-name(s);
RUN;
```

```
PROC PRINT DATA=input-table;
    FORMAT col-name(s) format;
RUN;
```

Format Name	Example Value	Format Applied	Formatted value
w.d	12345.67	5.	12346
w.d	12345.67	8.1	12345.7
COMMA*w.d*	12345.67	COMMA8.1	12,345.7
DOLLAR*w.d*	12345.67	DOLLAR10.2	$12,345.67
DOLLAR*w.d*	12345.67	DOLLAR10.	$12,346
YEN*w.d*	12345.67	YEN7.	¥12,346
EUROX*w.d*	12345.67	EUROX10.2	€12.345,67

Value	Format applied	Formatted value
21199	DATE7.	15JAN18
21199	DATE9.	15JAN2018
21199	MMDDYY10.	01/15/2018
21199	DDMMYY8.	15/01/18
21199	MONYY7.	JAN2018
21199	MONNAME.	January
21199	WEEKDATE.	Monday, January 15, 2018

- Week 4: prep data: data step
  - where, keep, drop, format

- functions: sum, mean, median, range(OF VAR1-VAR4), min, max, n, nmiss…
  upcase, lowcase, propcase, cats, substr(char, pos, <length>);
  month, year, day, weekday, qtr, today(), mdy(m,d,y), yrdif(d1,d2,'age')

- Week 5: analyze and report

  - TITLE<n> “”; FOOTNOTE<n> “”; ods noproctitle;

  - ods graphcis on;
    ods noproctitle;
    proc freq data=ds1 <order=[freq] nlevels noprint>;
        tables var1 var2 <row*col> / <nocum plots=freqplot(orient=horizontal
    scale=percent)><norow nocol nopercent crosslist list out=NAME>;
        format var2 monname.;
    run;

  - proc means data=ds1 <mean median min max maxdec=0 statistic(var)=var>;
        var var1;
        class groupvar1 groupvar2; *no sort needed;
        ways 0 1 2;
        output out=dsname mean=avevar;
    run;

- Week 6: SQL

  - proc sql;
    create table ds3 as
      select ds1.var1, ds2.var2, var3*2 as varx format=5.1., var4
        from ds1 [inner left outer right] join ds2
        on ds1.vary = ds2.varz
          <where expression1 and/or expres2>
            <order by var1 desc, var2>;
    drop table ds1; *delete tables;
    quit;