

# SAS Macros for Statistical Analysis and Data Management

Brandy R. Sinco  
Research Associate  
University of Michigan  
School of Social Work

## Outline

- Proc Univariate, T, F & Wilcoxon Test
- Normality Plot
- Proc Freq, Chi-Square Tests
- Linear and Logistical Regression
- Survey Procedures
- Building SAS Datasets from Multiple Files
- Assigning Macro Variables With %Let And Call Symput
- Debugging Macros
- Building a Macro Library

## Why Use Macros?

```
Proc Univariate Data=Data1; Var A B C; Run;
Proc Univariate Data=Data2; Var D E F; Run;
Proc Univariate Data=Data3; Var A B C; Run;
Proc Univariate Data=Data4; Var X Y Z; Run;
```

Macro Alternative:

```
%UniStat(A B C, Data1);
%UniStat(D E F, Data2);
%UniStat(A B C, Data3);
%UniStat(X Y Z, Data4);
```

## Univariate Summary Statistics

```
Data UNCLines;
  Format VARNAME DSNAME $30. N 7.0 MEAN MEDIAN STD SKEWNESS KURTOSIS
  MIN MAX 7.2 PROBT PROBS 7.4; Run;

%Macro UniStatN(VarName, DatSet);
/* Delete Previous Univariate Output File */
Proc Datasets MemType=Data; Delete UniOut UniOutNm; Run;

Proc Univariate Data=&DatSet NoPrint; Var &VarName;
  Output Out=UniOut N=N MEAN=MEAN MEDIAN=MEDIAN STD=STD MIN=MIN
  MAX=MAX SKEWNESS=SKEWNESS KURTOSIS=KURTOSIS PROBT=PROBT
  PROBS=PROBS;
Run;

/* Add Variable Name and Class Name to Univariate Output */
Data UniOutNm; Format VARNAME $30.; Set UniOut;
  VarName="&VARNAME"; DSNAME="&DATSET"; Run;

/* Append Most Recent Univariate Output to OutStats */
Data UNCLines; Set UniOutNm UNCLines; Run;
%MEnd UniStatN;
```

## Output Results to Comma Separated File Easily Imported Into Excel

```
DATA _NULL_; Set UNCLines; File unioutn;
  if _N_=1 then put
    "VARNAME" "," "N" "," "MEAN" "," "MEDIAN" "," "STD" "," "MIN" ","
    "MAX" "," "DSNAME" "," "PROBT" "," "PROBS";
  Put VARNAME "," N "," MEAN "," MEDIAN "," STD ","
  MIN "," MAX "," DSNAME "," PROBT "," PROBS; Run;
```

**Sample Table (Cholesterol Drug, Month 12 – Baseline):**

Variable	N	Mean	Median	Std Dev	p, T-Test	p, Wilcoxon
<b>M12 - BL</b>						
Total Chol	111	-10.1	-9	4.5	<.0001	<.0001
LDL Chol	109	-5.7	-5	2.2	.0003	.0002
HDL Chol	108	+4.1	+4	2.3	.0021	.0034

## Descriptive Statistics By Group

```
Data UCLines;
  Format VARNAME CLASSVAR DSNAME $30. CLSVALUE N 7.0 MEAN MEDIAN STD
  MIN MAX 7.2 PROBT PROBS 7.4; Run;

%Macro UniStat(VarName,ClassVar, DatSet);
/* Delete Previous Univariate Output Files */
Proc Datasets MemType=Data; Delete UniOut UniOutNm; Run;

Proc Sort Data=&DatSet; By &ClassVar; Run;

Proc Univariate Data=&DatSet NoPrint; Var &VarName; By &ClassVar;
  Output Out=UniOut N=N MEAN=MEAN MEDIAN=MEDIAN STD=STD MIN=MIN
  MAX=MAX PROBT=PROBT PROBS=PROBS; Run;

/* Add Variable Name and Class Name to Univariate Output */
Data UniOutNm; Format VARNAME CLASSVAR $30.; Set UniOut;
  VARNAME="&VARNAME"; /* VarName not in original Proc Univariate Output */
  CLASSVAR="&CLASSVAR"; DSNAME="&DATSET"; Rename
  &ClassVar=ClsValue; Run;

/* Append Most Recent Univariate Output to OutStats */
Data UCLines; Set UCLines UniOutNm; Run;
%MEnd UniStat;
```

## T-Test and Wilcoxon Test – Compare 2 Groups

```
%Macro TTestWil(OrigVar, TransVar, ClassVar, DataSet);

Proc Ttest Data=&DataSet;
  Class &ClassVar;
  Var &TransVar;
  title "T-Test: Response=&TransVar, Dataset=&DataSet";
  title2 "Comparison By: &ClassVar";
Run;

Proc NPar1Way Wilcoxon Data=&DataSet;
  Class &ClassVar;
  Var &OrigVar;
  Title "Wilcoxon: Response=&OrigVar, Dataset=&DataSet";
run;

%Mend TTestWil;
```

## One-Way ANOVA and Kruskal-Wallis Test Compare More Than 3 Groups

```
%Macro TestMore(OrigVar, TransVar, ClassVar, DataSet);

Proc GLM Data=&DataSet;
  Class &ClassVar;
  Var &TransVar;
  title "One-Way ANOVA Response=&TransVar, Dataset=&DataSet";
  title2 "Comparison By: &ClassVar";
  Model &TransVar = &ClassVar / Solution;
Run;

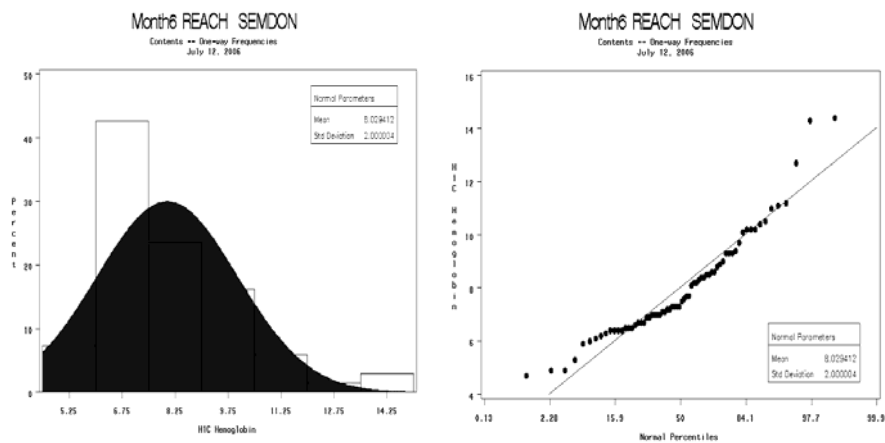
Proc NPar1Way Wilcoxon Data=&DataSet;
  Class &ClassVar;
  Var &OrigVar;
  Title "Wilcoxon: Response=&OrigVar, Dataset=&DataSet";
run;

%Mend TestMore;
```

## Normal Probability Plot And Histogram

```
%Macro UniGraph(Vary, DatSet);  
Proc Univariate Data=&DatSet;  
Var &VarY;  
Symbol1 V=Dot;  
  
Histogram / Normal(Mu=Est Sigma=Est Fill)  
    CFill=Blue Name="&VarName";  
Inset Mean Std / Header='Normal Parameters'  
    Position=(95,95) RefPoint=TR;  
  
QQPlot / Normal(Mu=Est Sigma=Est L=1 NoPrint)  
    PctlMinor PCTLSCALE Name="&VarName";  
Inset Mean Std / Header='Normal Parameters'  
    Position=(95,5) RefPoint=BR;  
Run;  
%Mend UniGraph;
```

## Normal Probability Plot And Histogram



## Proc Freq , Chi-Square Test

```
/* Create dataset for Proc Freq output */
Data FreqOut;
  Format VARNAME CLASSVAR $30. CLSVALUE VARVALUE 8.0 COUNT 8.2 DSET
    $30.; Run;

%Macro ChiFreq(VarName, ClassVar, DatSet);
/* Remove previous output file */
Proc Datasets Memtype=Data; Delete FreqOut0; Run;

Proc Freq Data=&DatSet;
  Tables &VarName * &ClassVar / ChiSq NoCum Out=FreqOut0 ; Run;

Data FreqOut0;
  Format VARNAME CLASSVAR DSET $30.;
  Set FreqOut0;
  Rename &VarName=VarValue &ClassVar=ClsValue;
  VARNAME="&VARNAME";
  CLASSVAR="&CLASSVAR";
  DSET="&DATSET"; Run;

Data FreqOut; Set FreqOut FreqOut0; Run;
%MEnd ChiFreq;
```

## Proc Freq Output

Variable	Control (N = 70)	Intervention (N = 68)	P Value
Caucasian, %	54.0	44.4	0.2850
Female, %	65.1	74.6	0.2442
High School Graduate, %	55.6	54.0	0.8579

## Proc Freq Other Options

**McNemar Chi-Square Test for comparison across timepoints.**

**Use "Agree" instead of "ChiSq".**

```
Tables &VarName * &ClassVar / AGREE NoCum Out=FreqOut0 ;
```

**Proc Freq Without Class Variable (I.E., Univariate Frequencies)**

```
/* Create dataset for Proc Freq output */
```

```
Data FreqOuN; Format VARNAME $30. VARVALUE COUNT 8.0 DSET $30.; Run;
```

```
%Macro ChiFreqN(VarName, DatSet);
```

```
/* Remove previous output file */
```

```
Proc Datasets Memtype=Data; Delete FreqOuN0; Run;
```

```
Proc Freq Data=&DatSet NoPrint;
```

```
Tables &VarName / ChiSq NoCum Out=FreqOuN0; Run;
```

```
Data FreqOuN0;
```

```
Format VARNAME DSET $30.; Set FreqOuN0;
```

```
Rename &VarName=VarValue;
```

```
VARNAME="&VARNAME";
```

```
DSET="&DATSET"; Run;
```

```
Data FreqOuN; Set FreqOuN FreqOuN0; Run;
```

```
%MEnd ChiFreqN;
```

## Linear Regression

```
/* Dataset to save paramter estimates from multiple runs */
```

```
Data RegAll; Format Dependent $30. Variable $30. DF Estimate StdErr tValue  
Probt 7.2; Run;
```

```
%Macro Regress(DepVar, Covar, ModelNum, DSNAME);
```

```
ods html; ods graphics on;
```

```
Proc datasets library=work memtype=data;
```

```
delete regout OutStat ResidStat ParamStat; run;
```

```
Title "Model# &ModelNum (&DepVar=&Covar)"; Title2 "Dataset=&DSNAME";
```

```
Proc Reg Data=&DSNAME; Model &DepVar=&Covar /
```

```
CP ADJRSQ R VIF INFLUENCE;
```

```
Output Out=RegOut p=yhat student=stdresid U95=U95 L95=L95
```

```
h=infl press=press cookd=cookd;
```

```
ODS Output OutputStatistics=OutStat ResidualStatistics=ResidStat
```

```
ParameterEstimates=ParamStat;
```

```
Run;
```

```
/* Add new output to RegOut */
```

```
Data RegAll; Set RegAll ParamStat;
```

```
Keep Dependent Variable DF Estimate StdErr tValue Probt; Run;
```

```
ods graphics off; ods html close;
```

```
%Mend Regress;
```

## Stepwise Regression

```
%macro stepreg(DepVar, covar,modelNum,selectyp,slentry,slstay, DSNAME);
Proc datasets library=work memtype=data;
delete regout; /* delete output from previous run */
run;

Title "Model# &ModelNum (&DepVar=&Covar)";
Title2 "Dataset=&DSNAME";
Proc Reg Data=&DSNAME;
  Model &DepVar=&Covar / selection=&selectyp sle=&slentry slstay=&slstay
  CP ADJRSQ R VIF;
  Output Out=RegOut p=yhat student=stdresid U95=U95 L95=L95
  h=infl press=press;
Run;
%mend stepreg;
```

## Logistical Regression

```
Data LogitRegOut; Format DSNAME $30. OUTCOME $30.; Run;

%Macro LogitReg(Outcome, Predict, Covar, DatSet); ods html; ods graphics on;

/* Delete previous output datasets */
Proc Datasets MemType=Data; Delete LogitRegOut0 LogitRegOut1; Run;

Proc Logistic Data=&DatSet Descending;
  Model &OutCome = &Predict &Covar / CLODDS=BOTH influence iplots;
  Title "Logistic Model: Outcome: &OutCome; Predictor: &Predict";
  Title2 "Covariates: &Covar";
ODS Output ParameterEstimates=LogitRegOut0; Run;

Data LogitRegOut1;
Format DSNAME $30. OUTCOME $30.;
Set LogitRegOut0;
DSNAME="&DatSet";
OUTCOME="&OutCome"; Run;

Data LogitRegOut; Set LogitRegOut LogitRegOut1; Run;
ods graphics off; ods html close; %Mend LogitReg;
```



## Repeated Measures With Proc Mixed

```
%Macro MixExplo(ClassVar, FixedVar, CorrType, Method, Dataset);
ods html; ods graphics on;

/* Delete previous residual file before running Proc Mixed */
Proc datasets lib=work;
  delete resids SolutionR SolutionF influence;   Run;

Title "Mixed Model: Corr=&CorrType, Covar=&FixedVar";
Proc Mixed Method=&Method Data=&Dataset NoCIPrint Covtest;
Class &ClassVar;
Model A1c = &FixedVar /Solution Influence(effect=ID Est) outp=resids;
Random Intercept / Subject=SiteID Solution;
Repeated TimepointN / Subject=ID Type=&CorrType;
ODS Output SolutionR=SolutionR SolutionF=SolutionF Influence=Influence;   Run;

ods graphics off; ods html close;
%MEnd MixExplo;
```

## Repeated Measures With Proc GenMod

```
/* Create dataset to store multiple runs of Proc GenMod */
Data GEEBinAll;
Format DSET Dependent Variable Parameter Level1 $30.;
Format Estimate StdErr ChiSq ProbChiSq 7.2;   Run;

/* Categorical Repeated Measures */
%Macro GEEBin(YVar, ClassVar, VarList, DSET, VarType);
Proc datasets library=work memtype=data;
/* delete output from previous run */ delete GEEBinOut GEERegBin; Run;

Title "&YVar, VarType=&VarType";
Proc GenMod Data=&DSET Descending;
Class REACHID &ClassVar;
Model &YVar=&VarList &ClassVar / dist=bin;
Repeated Subject=REACHID / type=&VarType modelse;
ODS OUTPUT GEEModPEst=GEERegBin; Run;

Data GEERegBin;   Set GEERegBin;
Dependent="&YVar"; Variable="&VarList";
DSET="&DSET";   Run;

/* Add new output to RegOut */
Data GEEBinAll; Set GEEBinAll GEERegBin; Run;
%MEnd GEEBin;
```

## Survey Data: Proc SurveyFreq

```
/* Add variable name, class variable ,data set name to Proc Survey Freq output
*/
Data FreqOut; Format VARNAME CLASSVAR DSET $30.; Run;

%Macro SurvFreq(VarName, ClassVar, DatSet, StratVar, WtVar, ClusVar);
/* Remove previous output file */
Proc Datasets Memtype=Data; Delete FreqOut0; Run;

Proc SurveyFreq Data=&DatSet;
Tables &ClassVar*&VarName / CL ROW WCHISQ CHISQ;
strata &stratvar; weight &wtvar; cluster &clusvar;
ODS Output CrossTabs=FreqOut0; Run;

Data FreqOut0;
  Format VARNAME CLASSVAR $30.;
  Set FreqOut0;
  Rename &VarName=VarValue &ClassVar=ClsValue;
  VARNAME="&VARNAME";
  CLASSVAR="&CLASSVAR";
  DSET="&DATSET"; Run;

Data FreqOut;
  Set FreqOut FreqOut0; Run;
%MEnd SurvFreq;
```

## Survey Data: Proc SurveyLogistic

```
Data Stat_All;
  Format OutName $20. PredVars $40. DSNAME $20.; Run;

%Macro SurvLgt(Outcome, Predict, DatSet, StratVar, ClusVar, WtVar, ClassVar);
Proc Datasets;
delete Stat0; Run;

Title "Outcome=&Outcome, Dataset=&DatSet";
Title2 "Predictors=&Predict";
Proc SurveyLogistic Data=&DatSet;
Weight &WtVar; Strata &StratVar; Cluster &ClusVar; Class &ClassVar;
model &Outcome (desc)=&Predict / CLODDS;
ods output ParameterEstimates=Stat0; run;

Data Stat1;
  Set Stat0;
  Format OutName $20. PredVars $40. DSNAME $20.;
  OutName="&Outcome";
  PredVars="&Predict";
  DSNAME="&DatSet"; Run;

Data Stat_All;
Set Stat_All Stat1; Run;
%MEnd SurvLgt;
```

### Survey Data: Proc SurveyRegress

```

Data Reg_All;
  Format OutName $20. PredVars $40. DSNAME $20.; Run;

%Macro SurvReg(Outcome, Predict, DatSet, StratVar, ClusVar, WtVar, ClassVar);
Proc Datasets;
delete Reg0; Run;

Title "Outcome=&Outcome, Dataset=&DatSet";
Title2 "Predictors=&Predict";
Proc SurveyReg Data=&DatSet;
Weight &WtVar; Strata &StratVar; Cluster &ClusVar; Class &ClassVar;
model &Outcome=&Predict / SOLUTION;
ods output ParameterEstimates=Reg0; Run;

Data Reg1;
  Set Reg0;
  Format OutName $20. PredVars $40. DSNAME $20.;
  OutName="&Outcome";
  PredVars="&Predict";
  DSNAME="&DatSet"; Run;

Data Reg_All;
Set Reg_All Reg1; Run;
%MEnd SurvReg;

```

### Sample Logistical Regression Output

Variable	Predictors	Dataset	95%CI	Prob ChiSq
<b>agecat2</b>	agecat2 agecat3 genderind southwest educ	AAAll	0.60 (0.44, 0.82)	0.0012
<b>agecat3</b>	agecat2 agecat3 genderind southwest educ	AAAll	0.35 (0.24, 0.49)	<0.0001
<b>genderind</b>	agecat2 agecat3 genderind southwest educ	AAAll	1.65 (1.25, 2.19)	0.0005

## Building SAS Datasets from Multiple Files

```
/* Read raw data */
%Macro Raw(DatSet,InData);
data &DatSet;

    infile "N:\Brandy\SASData\Baseline\&InData";
    input
#1 @3 IDNUM1 $8.    @11 IWRCODE 2.    @13 IWMONTH 2.
    @15 IWDAY 2.    @17 IWYEAR 4.    @21 NUMMIN 2.
    @23 RECRUIT 1.    @24 DOBMONTH 2.
    @26 DOBDAY 2.    @28 DOBYEAR 4.

#2 @11 BLV36 1.    @12 BLV37 1.    @13 BLV38 1.
Run;
%mend raw;

%raw(TEMPDAT1,Care0724.dat);
%raw(TEMPDAT2,Baseline2.txt);
%raw(TEMPDAT3,base0708.txt);

/* Combine the Input Datasets */
Data TempDatA;
    set Tempdat1 Tempdat2 Tempdat3;
Run;
```

## Assigning Macro Variables: %Let and Call Symput

**%LET.** This SAS system macro, &SYSDATE, with a more desirable format.  
options symbolgen; **%Global Today;**

```
%Macro CompTDay;
%Let Month0=%SUBSTR(&SYSDATE,3,3);
%Let Day0=%SUBSTR(&SYSDATE,1,2);    %Let Year0=%SUBSTR(&SYSDATE,6,2);
%IF &MONTH0=JAN %THEN %LET MONTHB=January;
%IF &MONTH0=FEB %THEN %LET MONTHB=February;
%IF &MONTH0=MAR %THEN %LET MONTHB=March;
%IF &MONTH0=APR %THEN %LET MONTHB=April;
%IF &MONTH0=MAY %THEN %LET MONTHB=May;
%IF &MONTH0=JUN %THEN %LET MONTHB=June;
%IF &MONTH0=JUL %THEN %LET MONTHB=July;
%IF &MONTH0=AUG %THEN %LET MONTHB=August;
%IF &MONTH0=SEP %THEN %LET MONTHB=September;
%IF &MONTH0=OCT %THEN %LET MONTHB=October;
%IF &MONTH0=NOV %THEN %LET MONTHB=November;
%IF &MONTH0=DEC %THEN %LET MONTHB=December;
%LET TODAY=&MonthB &Day0, 20&Year0;
%Mend CompTDay;
```

**NOTE:** The %GLOBAL statement enables the macro variable, &TODAY, to be recognized outside of the macro where it was created.

## Call Symput Example: 1, 2, 5 Range With Proc GChart

```
%Macro BarChart(SelectX, DSNAME);
Proc Univariate Data=&DSNAME;  Var &SelectX;
  Output Out=UniDatX MAX=XMAX MIN=XMIN RANGE=XRANGE;  Run;

Data CompBinX;  Set UniDatX;
/* Define X Scale */
If 0 < Xrange <= 1 then XNewRan=1;  If 1 < Xrange <= 2 then XNewRan=2;
If 2 < Xrange <= 5 then XNewRan=5;  If 5 < Xrange <=10 then XNewRan=10;
If 10 < Xrange <= 20 then XNewRan=20;  If 20 < Xrange <= 50 then XNewRan=50;

If 50 < Xrange <= 100 then XNewRan=100; X_Inc=XNewRan/10;
If (Xmin < ((Round(Xmin/X_Inc) - .5) * X_Inc))
Then  X_Low = (Round(Xmin/X_Inc)*X_Inc) - X_Inc;
Else X_Low = (Round(Xmin/X_Inc)*X_Inc;
If (Xmax > ((Round(Xmax/X_Inc) + .5) * X_Inc))
Then  X_High = (Round(Xmax/X_Inc)*X_Inc) + X_Inc;
Else X_High = (Round(Xmax/X_Inc)*X_Inc;

Call Symput('XLOW', X_Low); Call Symput('XHIGH', X_High);
Call Symput('XINC', X_Inc); Run;

Proc Gchart Data=&DSNAME;
VBAR &SELECTX / MAXIS = &XLOW TO &XHIGH BY &XINC TYPE=PERCENT; Run;
%Mend BarChart;
```

## Debugging Macros: SymbolGen and MPrint

- **Symbolgen:** will cause the SAS Log to print a line, telling the programmer how each macro variable resolved.
- Options Symbolgen;
- In SAS log, &SELECTX resolved to TORQUE.
- **MPrint:** When used on the Options statement, will cause the SAS Log to display each line of code that is generated by a macro.
- **Use Macro to Prevent Code From Running – Alternative to “commenting out” code.**

```
%Macro DontRun;
Proc Freq Data=TestData;
Tables Var1*Var2;
Run;
%Mend DontRun;
```

### **Building a Macro Library - %Include**

- When tasks, such as comparing 2 numeric variables by a class, running chi-square frequency tests, regression, etc. will be performed many times, an analyst can save time and energy by storing macros in a directory. When the macro needs to be used in a program, it can be easily included by the %INCLUDE statement.
- %INCLUDE
- 'N:\brandy\SASMacro\Compare2.sas';
- %TtestWil(BrakePr, SqrtBrPr, Car, BaseData);
- %TtestWil(BrakeTem, LogTem, Car, BaseData);

### **Contact Information**

Brandy R. Sinco  
Research Associate  
University of Michigan School of Social Work  
1080 S. University St.  
Box 183  
Ann Arbor, MI 48109-1106

Phone: 734-763-7784  
E-Mail: [brsinco@umich.edu](mailto:brsinco@umich.edu)

Diabetes Work Funded by Centers for Disease Control and Prevention,  
Racial and Ethnic Approaches to Community Health (REACH) 2010,  
Grant Number: U50/CCU522189-02.