Using SAS to Analyze Longitudinal Study

Abbas S. Tavakoli, DrPH, MPH, ME; Kirby Jackson, MS, PhD(c); Linda Moneyham, DNS, RN, FAAN; & Carolyn Murdaugh, PhD, RN, FAAN

**Background:** The importance of choosing what type of statistical program that we use to analyze the longitudinal study are growing as a specialty due to the fact computerized data analyses has become the basic for scientific research. SAS is the most powerful statistical program to analyze longitudinal study. All data analysis tasks can almost be written into a program file that serves as written documentation of what you did. Within the data analysis program, use comments to describe what each step does and include TITLE statements that will print a brief description of the analysis and a date on your printed output. This simple task will save time trying to guess the type of analysis and its creation date down the road. Procedural steps perform the actual data analyses which include descriptive and inferential statistics. The program itself should consist of combinations of these steps structured in a meaningful way.

**Aim:** The purpose of this poster is to give an example of SAS programming set up to analyze the longitudinal study.

**Conclusion:** SAS is the most powerful statistical program in data analyses for longitudinal study.

**Keywords:** BASE STAT FUNCTIONS

University of South Carolina, College of Nursing.
Introduction

The programming of longitudinal data sets is, as required for intervention research, poses a challenge to new investigators. Every programmer goals should include writing clear and efficient codes. SAS provides practical and efficient way to analyze longitudinal and complex design.

Purpose

The purpose of this poster is to give an example of SAS programming set up to analyze the longitudinal study.

Background

The longitudinal data set was generated in a study which 420 participants were recruited to test a peer-based social support intervention designed for a population of rural women with HIV disease in Alabama, Georgia, North Carolina, and South Carolina. The basic structure for this study is that of a repeated measures analysis with four measures taken per subject and subjects randomly assigned: 1) a control group that receives the usual care provided by the community-based organization; 2) a group who receives the peer counseling intervention via face-to-face meetings with the peer counselor; and 3) a group who receives the peer counseling intervention via telephone.

Conclusion

The statistical software (SAS) is an integrated software package for data management, analysis, and reporting. Descriptive and inferential statistics programs can be written using SAS. SAS should be selected to analyze the data for two important reasons: 1) SAS is a very flexible package that can accommodate a very large number of variables. 2) SAS allows control over statistical modeling algorithms (See examples of SAS program for this project).
PROGRAM1. Part of Program RWHPii1PRM.SAS

option nodate nocenter;
libname library 'C:\abbast\moneyham\rwhpiil\';
libname rwhpiil 'c:\abbast\moneyham\rwhpiil\';

**** this program written for southern women health survey ****;
**** Phase II ****;
**** filename : rwhpiilprm.sas ****;
**** author : abbas tavakoli ****;
**** last modification : 05/22/08 ****;

*** Part of format ***;
PROC FORMAT library=rwhpiil.formats;

VALUE wSTATEF
  1 = 'SOUTH CAROLINA'
  2 = 'GEORGIA'
  3 = 'ALABAMA'
  4 = 'north carolina'
  97 = 'do not know'
  98 = 'refused'
  99 = 'does not apply';

VALUE LCITYF
  1 = 'NO'
  2 = 'YES'
  7 = 'do not know'
  8 = 'refused'
  9 = 'does not apply';

VALUE groupf
  1 = 'control'
  2 = 'face to face'
  3 = 'telephone'
  7 = 'do not know'
  8 = 'refused'
  9 = 'does not apply';

data one;
  set rwhpiil.rwhpiil;

data rwhpiil.rwhpiils;
  set one;

*** Part of label ***;
LABEL
  DATE = 'DATE'
  PID = 'PARTICIPANT ID#'
  IID = 'INTERVIEWER ID#'
  DOB = 'DATE OF BIRTH'
  STATE = 'LIVING in sc'
  Wstate = 'what state do you live';

*** part of format statement ***;
FORMAT
PROGRAM2. Part of Program RWHPii2PRM.SAS

option nodate nocenter;

libname library 'C:\abast\moneyham\rwhpii2\';
libname rwhpii2 'C:\abast\moneyham\rwhpii2\';

*** Part of format ***;
PROC FORMAT library=rwhpii2.formats;

VALUE bLCITYF 1 = 'NO'
2 = 'YES'
7 = 'do not know'
8 = 'refused'
9 = 'does not apply'
;
VALUE bhivstf 1 = 'hiv asymptomatic'
2 = 'hiv symptomatic'
3 = 'aids'
7 = 'do not know'
8 = 'refused'
9 = 'does not apply'
;
VALUE bgroupf 1 = 'control'
2 = 'face to face'
3 = 'telephone'
7 = 'do not know'
8 = 'refused'
9 = 'does not apply'
;
data one;
  set rwhpii2.rwhpii2;

data rwhpii2.rwhpii2s;
  set one;

*** Part of label ***;
LABEL
  bDATE = 'DATE'
bPID = 'PARTICIPANT ID#'
biID = 'INTERVIEWER ID#'
bMARITAL = 'MARITAL STATUS'
blIVING = 'CURRENT LIVING SITUATION'
;

*** part of format statement ***;
FORMAT
PROGRAM3. Part of Program RWHPii3PRM.SAS

option nodate nocenter;

libname library 'C:\abast\moneyham\rwhpii3\';
libname rwhpii3 'c:\abast\moneyham\rwhpii3\';

**** this program written for southern women health survey ****;
**** Phase II ****;
**** filename : rwhpii3prm.sas ****;
**** author   : abbas tavakoli ****;
**** last modification : 05/22/08 ****;

*** Part of format ***;
PROC FORMAT library=rwhpii3.formats;

VALUE cLCITYF 1 = 'NO'
               2 = 'YES'
               7 = 'do not know'
               8 = 'refused'
               9 = 'does not apply'
;
value chivstf 1='hiv asymptomatic'
               2='hiv symptomatic'
               3='aids'
               7 = 'do not know'
               8 = 'refused'
               9 = 'does not apply'
;
VALUE cgroupf 1 = 'control'
               2 = 'face to face'
               3 = 'telephone'
               7 = 'do not know'
               8 = 'refused'
               9 = 'does not apply'
;
DATA one;
   set rwhpii3.rwhpii3;

DATA rwhpii3.rwhpii3s;
   set;

*** Part of label ***;
LABEL
cDATE  = 'DATE'
cPID   = 'PARTICIPANT ID#'
cIID   = 'INTERVIEWER ID#'
cMARITAL = 'MARITAL STATUS'
cLIVING = 'CURRENT LIVING SITUATION'
;
*** part of format statement ***;
FORMAT

PROC FORMAT library=rwhpii4.formats;

VALUE dLCITYF 1 = 'NO'
               2 = 'YES'
               7 = 'do not know'
               8 = 'refused'
               9 = 'does not apply'
;
VALUE dhivstf 1 = 'hiv asymptomatic'
               2 = 'hiv symptomatic'
               3 = 'aids'
               7 = 'do not know'
               8 = 'refused'
               9 = 'does not apply'
;
VALUE dgroupf 1 = 'control'
               2 = 'face to face'
               3 = 'telephone'
               7 = 'do not know'
               8 = 'refused'
               9 = 'does not apply'
;
*** Part of label ***;
LABEL
dDATE   = 'DATE'
dPID    = 'PARTICIPANT ID#'
dIID    = 'INTERVIEWER ID#'
dMARITAL = 'MARITAL STATUS'
dLIVING = 'CURRENT LIVING SITUATION'

*** part of format statement ***;
FORMAT
dlcity dlcityf. dgroup dgroupf. dhivst dhivstf.
;
run;

PROGRAM5. Part of Program RWHPii1dis.SAS

option nodate nocenter yearcutoff=1910;;

libname library 'c:\abbast\moneyham\rwhpiii1\';
libname rwhpii1 'c:\abbast\moneyham\rwhpiii1\';

**** this program written for southern women health survey ****;
*** Phase II ****;
**** filename : rwhpiii1dis.sas ****;
**** author : abbas tavakoli ****;
**** last modification : 05/22/08 ****;

** part of format **;
proc format ;
  VALUE martgF 1 = 'SINGLE/sep/div/wid'
                 2 = 'MARRIED/living/other'
;
  value ylearngf 1='2001'
                  2='1996-2000'
                  3='1991-1995'
                  4='>1990';

  value livinggf 1='alone'
                  2='others';

  value educgf 1='less hs'
               2='hs'
               3='college/more'
;
data one;
set rwhpiii1.rwhpiii1s;

*** part of program to set missing and reverse coding **;
ARRAY items  race marital living  ;
do over items;
  if items=9 then items=.;
end;

array itema time2 time5 time6  sas2 sas4 sas6 fam1-fam5  ;
do over itema;
  itema = 5 - itema;
end;

**** Part of syntax to create scales and subscales ****;
tfreq  = sum (of freq1-freq31);
tboth  = sum (of both1-both31);
tphys  = sum (of phys1-phys11);
ttime  = sum (of timel-time6 );

**** Part of syntax to create new variables ****;
if income =1 then incomeg=1;
    else if income=2 then incomeg=2;
    else if 2<income<7 then incomeg=3;

if yrtoldg=1 or yrtoldg=2 then ytoldgg=1;
    else if yrtoldg=3 or yrtoldg=4 then ytoldgg=2;

*** part of label for new varaibles ***;
label
tfreq  = 'total /hivsymfreq '
tboth  = 'total / hivsymboth '
tphys  = 'total / fsq activity of daily living'
ttime  = 'total / fsq work function '
tss    = 'total / social support'
;
*** part of format statement ***;
format  martg martgf. ylearng yrtoldg ylearngf.
;
** Part of Syntax for Descriptive Statistics ***;
ods html file="C:\"abbast\moneyham\rwhpii1\freq.html" ;
proc freq data=one;
tables id -- bar21;
    title ' frequency tables ';
    title2 'rural women's health project';
run;
proc means data=two;
var ages nchild child18 hourwk ;
title ' means ';
    title2 'rural women's health project';
run;
ods html close;
run;

PROGRAM6. Part of program to Check the Data

proc format;
    value $missf ' ' = 'missing'
        other= 'nonmissing';

**** Macro that takes lower and upper limits for numeric variables and*;
**** an id variable to print out an exception report to the output
Windows ***;

%macro range (dsn,var,low,high,idvar);
data _null_; set &dsn ; file print;
    if (&var lt &low and &var ne .) or &var gt &high then
        put "&idvar:" &idvar @15 " variable: &var"
            @35 "value:" &var
            @50 "out-of-range";
run;
%mend range;

%range (one,cope1,0,3,id);
**** checking data: counting missing values  ***;
proc freq;
   tables _character_ / nocum missing;
   format _character_ $missf.;
   title 'frequency tables /missing';
   title2 'rural women's health project II';
run;

proc means nmiss;
   title 'means /missing';
   title2 'rural women's health project II';
run;

PROGRAM 7. Part of program RWHPiiMERGEOUT.SAS
option nodate nocenter yearcutoff=1910;;

libname library 'c:\abbast\moneyham\rwhpii1'';
libname rwhpii1 'c:\abbast\moneyham\rwhpii1'';

**** this program written for southern women health survey  ****;
**** filename : rwhpiimergeout.sas  ****;
**** author : abbas tavakoli  ****;
**** last modification : 05/22/08  ****;
data one;
set rwhpii1.rwhpii1s;
data two;
   set one;
   .
proc sort;
   by id;

libname library 'c:\abbast\moneyham\rwhpii2'';
libname rwhpii2 'c:\abbast\moneyham\rwhpii2'';

proc format ;
   value btmastgf 0='no abuse' 1='abuse';
data three;
set rwhpii2.rwhpii2;
data four;
   set three;
   id = bid;
   .
proc sort;
   by id;
libname library 'c:\abbast\moneyham\rwhpii3\';
libname rwhpii3 'c:\abbast\moneyham\rwhpii3\';

data five;
set rwhpii3.rwhpii3;

data six;
  set five;
id = cid;
  .
  .
proc sort;
  by id;

libname library 'c:\abbast\moneyham\rwhpii4\';
libname rwhpii4 'c:\abbast\moneyham\rwhpii4\';

data seven;
set rwhpii4.rwhpii4;

data eight;
  set seven;
id = did;
  .
  .
proc sort;
  by id;

proc sort;
  by id;

libname rwhpiilf  'c:\abbast\moneyham\rwhpiil';
libname rwhpii2f   'c:\abbast\moneyham\rwhpii2';
libname rwhpii3f   'c:\abbast\moneyham\rwhpii3';
libname rwhpii4f   'c:\abbast\moneyham\rwhpii4';

options fmtsearch=(rwhpiilf rwhpii2f rwhpii3f rwhpii4f);

data final;
  merge two (in=a) four (in=b) six (in=c) eight (in=d);
  by id;

libname rwhpimrg 'c:\abbast\moneyham\merge2';

data rwhpimrg.rwhpimrg;
  set final;
run;
PROGRAM8. Part of program RWHPiiMERGEDIS.SAS

option nodate nocenter yearcutoff=1910;;

**** this program written for southern women health survey ****;
**** filename : rwhpiimergedis.sas ****;
**** author : abbas tavakoli ****;
**** last modification : 05/22/08 ****;

libname rwhipmrg 'c:\abbast\moneyham\merge2';
libname rwhpii1f 'c:\abbast\moneyham\rwhpii1';
libname rwhpii2f 'c:\abbast\moneyham\rwhpii2';
libname rwhpii3f 'c:\abbast\moneyham\rwhpii3';
libname rwhpii4f 'c:\abbast\moneyham\rwhpii4';
options fmtsearch=(rwhpii1f rwhpii2f rwhpii3f rwhpii4f);

proc format ;
   value cesdgf 1=' less 16'
          2='16 +';
   value cesdgbf 1=' less 16'
                 2='16 -25'
                 3='26-35'
                 4='36+';
   value groupgf 1='control'
                 2='intervention';
data one;
   set rwhpimrg.rwhpimrg;
   by pid;
   if last.pid;

data two;
   set one;
   if 0<=tcesd<16 then cesdg=1;
   else if 15<tcasd<61 then cesdg=2;
   if 0<=btcesd<16 then bcesdg=1;
   else if 15<btcesd<61 then bcesdg=2;
   if group=1 then groupg=1;
   else if group=2 or group=3 then groupg=2;
   if cgroup=1 then cgroupg=1;
   else if cgroup=2 or cgroup=3 then cgroupg=2;
   format cesd bcesd cesdgf cesdgb bcesdgb cesdgbf groupg cgroupg groupgf .;
proc freq data=two;
   Tables pid  group bgroup cgroup groupg cgroupg;
   title ' frequency tables ';
   title2 'rural women"s health project';
   title3 'merge file ';
**Program 9. Example of SAS Procedure**

**** descriptive statistics ****;
proc freq data=one;
    tables id -- dsas11;
    title 'frequency tables';
    title2 'rural women"s health project II';
run;
proc means data=two;
    var ages nchild child18 hourwk income actual freq1 -- phys11;
    title 'means';
    title2 'rural women"s health project II';
run;
**** Inferential Statistics ****;
%macro corr (q);
    proc corr nocorr alpha nomiss data=two;
        var &q;
    ;
    title 'Reliability coefficient';
    title2 'rural women"s health project II';
%mend corr;
%corr (freq1-freq31);
run;
proc corr data=two;
    var tfreq -- tcopeis tcesd --;
    title 'correlation coefficient';
    title2 'rural women"s health project II';
run;
%macro glm (d,i);
    proc glm data=two;
        class &i;
        model &d = &i;
        lsmeans &i / pdiff cl adjust=tukey;
    ;
    title 'glm';
    title2 'baseline';
    title3 'rural women"s health project II';
%mend glm;
%glm (tcesd tlslp tlsln tlslo tls1f treas tsctrl tsas,state);
ods html file="C:\abbast\moneyham\rwhpii1\rwhpii1reg.html"
style=fancyprinter;
%macro reg (d,i,t);
    proc reg data=two;
        model &d = &i / stb pcorr2 scorr2;
    ;
    title 'Regression model' &t;
title2 'rural women''s health project II';

%mend reg;

%reg (tls1p,tistig, present life satisfaction );
run;
ods html close;
run;
ods html file="C:\abbast\moneyham\rwhpii\rwhpii1fact.html" style=fancyprinter;
%macro fact (q,n,t);
proc factor data=two method=prin priors=smc scree rotate=promax reorder
 flag=.35 nfact=&n msa ;
 var &q ;
title ' factor analysis ' &t;
title2 'rural women''s health project II';
%mend fact;

%fact (stig1-stig12,2, Stigma scale two factor );
run;
ods html close;
run;
*******************************************************;
**** mix model                                      ***;
******
*************************************************;
%macro mix (a,b,c,d,e);
proc mixed noclprint noitprint;
class id time &a;
 model &b = &c / s ;
 repeated time / type=&d sub=id(group) ;
 title ' mixed model ' &e ;
title2 'rural women''s health project II';
%mend mix;

%mix (group site,cesd,group time group*time site site*time ,cs,reduced model/cs);
run;

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration. Other brand and product names are trademarks of their respective companies.