Brief introduction of SAS

First, double-click on the "SAS" icon on the desktop to open up SAS. You should find four types of windows. They are:

1. Explorer/Results Window: a tool for browsing SAS libraries, program files, and the results of statistical procedures.
2. Program editor Window: create, edit, and execute SAS programs.
3. Output Window: output requested by the SAS program is printed here after it is submitted.
4. Log Window: a tool for diagnosing problems with SAS programs. It is a good habit to check the Log Window after submitting a SAS program in order to ensure that the program did not encounter any errors.

In addition to these windows, other windows are displayed under certain operations such as graph and help window. Also pay attention to the tool bar at the top the screen. This contains several commonly used SAS operations such as:

a. Create a file.
b. Open a file in a new program editor window.
c. Save a file.
d. Print editor contents.
e. Display contents of the editor as they appear when printed.
f. Some edit operations: cut, copy, paste and undo.
g. Add a new library.
h. Open the SAS Explorer.
i. Submit a SAS Program.
j. Clear all.
k. Interrupt the executing program.
l. Get help on the SAS System.

SAS program contains the following two major parts:
DATA step: Starts from the keyword DATA; create and manipulate SAS datasets.
PROC step – Starts with the keyword PROC followed by the name of the procedure; analyze the SAS dataset created in DATA step, and produce results or report.

When we get a dataset, we always explore it first to understand what all the variables are and to make sure that everything is fine. The commonly used procedures are, *PRINT, CONTENTS, MEANS, FREQ* etc.:
PROC PRINT: displays the values of the variables in a SAS data set.
PROC CONTENTS: displays the descriptions of the contents of a SAS data set.
PROC MEANS: produces simple univariate descriptive statistics for numeric variables.
PROC FREQ: produces 1-way to **n**-way frequency and crosstabulation tables.

Commonly used SAS procedures for LDA:

1. PROC GLM

The GLM procedure uses the method of least squares to fit general linear models for the continuous outcome data. It can perform regression, analysis of variance, analysis of covariance, multivariate analysis of variance, and partial correlation.

2. PROC LOGISTIC

The LOGISTIC procedure fits linear logistic regression models for binary response data by the method of maximum likelihood.

3. PROC CATMOD

The CATMOD procedure provides a wide variety of categorical data analyses, such as linear models to functions of response frequencies, log-linear modeling, logistic regression, and repeated measurement analysis etc.

4. PROC GENMOD

The GENMOD procedure can fit models to correlated responses by the GEE method. The GENMOD procedure fits generalized linear models that is an extension of traditional linear models that allows the mean of a population to depend on a linear predictor through a nonlinear link function and allows the response probability distribution to be any member of an exponential family of distributions.

5. PROC MIXED

The MIXED procedure fits random-effects models and a variety of mixed linear models to data and enables you to use these fitted models to make statistical inferences about the data. A mixed linear model is a generalization of the standard linear model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and nonconstant variability.

An Introduction to SAS

April 16, 2003

Getting Started

There are two kinds of steps in SAS programs.

- `DATA` step: tells SAS about the data, reads data from a file
- `PROC` step: tells SAS what statistical analyses to perform and provides specifications for those analyses

How to read in data from a .raw file:

```
data somename;
     infile 'filename.raw';
     input varlist;
;
run;
```

When giving the filename you must specify the directory of the file:

```
     infile 'C:\My Documents\My SAS Files\file.raw';
     infile '/ga/jager/mysas/file.raw';
```

Some useful `proc` steps:

- `proc contents` (In Stata: `describe`)
- `proc means` (In Stata: `summarize`)
- `proc prints` (In Stata: `list`)

The SAS commands should be entered into the SAS program editor window. Select 'Submit' under the 'Run' menu to perform the `data` or `proc` step. Your output will appear in another window - the SAS output window.

Example

Suppose for the back data I want to fit the model

$$\texttt{alert}_{ij} = \beta_0 + \beta_1 I(\texttt{group 1})_i + \beta_2 \texttt{time}_{ij} + \beta_3 I(\texttt{group 1})_i * \texttt{time}_{ij} + \epsilon_{ij}$$

assuming an exponential correlation structure.

We will begin by doing our "data manipulation" in Stata.

```
. infile id group painvrs1 painvas1 anxvas1 altvas1 t1 painvrs2 painvas2
anxvas2 altvas2 t2 painvrs3 painvas3 anxvas3 altvas3 t3 painvrs4
painvas4 anxvas4 altvas4 t4 using back.raw

. reshape long painvrs painvas anxvas altvas t, i(id)

. replace id=. if(id==-9)
. replace group=. if(group==-9)
. replace painvrs=. if(painvrs==-9)
. replace painvas=. if(painvas==-9)
. replace anxvas=. if(anxvas==-9)
. replace altvas=. if(altvas==-9)
. replace t=. if(t==-9)

. gen gr1=1 if group==1
. replace gr1=0 if group!=1
. gen gr2=1 if group==2
. replace gr2=0 if group!=2
. gen gr1time=gr1*t
. gen gr2time=gr2*t

. outfile id _j group painvrs painvas anxvas altvas t gr1 gr2 gr1time gr2time using back
```

Now we can read `backlong2.raw` into SAS:

```
data back;
     infile '/aa/jager/hs333/backlong2.raw';
     input id j group painvrs painvas anxvas altvas t gr1 gr2 gr1time gr2time;
;
run;
```

We can get some summary statistics and look at the data:

```
proc contents data=back;
run;
```

```
proc means data=back;
run;
proc print data=back(obs=5);
run;
```

Output:

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
7	altvas	Num	8	48
6	anxvas	Num	8	40
9	gr1	Num	8	64
10	gr2	Num	8	72
11	gr1time	Num	8	80
12	gr2time	Num	8	88
3	group	Num	8	16
1	id	Num	8	0
2	j	Num	8	8
5	painvas	Num	8	32
4	painvrs	Num	8	24

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
8	t	Num	8	56

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
id	108	14.0000000	7.8251930	1.0000000	27.0000000
j	108	2.5000000	1.1232463	1.0000000	4.0000000
group	108	1.4814815	0.5019864	1.0000000	2.0000000
painvrs	107	2.2149533	0.8244843	1.0000000	5.0000000
painvas	106	29.0283019	20.8379039	0	86.0000000
anxvas	104	26.4615385	23.9534026	0	99.0000000
altvas	105	54.8666667	27.2340520	2.0000000	97.0000000
t	107	443.6728972	270.1152366	25.0000000	890.0000000
gr1	108	0.5185185	0.5019864	0	1.0000000
gr2	108	0.4814815	0.5019864	0	1.0000000
gr1time	107	232.5700935	296.6862936	0	870.0000000
gr2time	107	211.1028037	289.9279655	0	890.0000000

Obs id j group painvrs painvas anxvas altvas t gr1 gr2 gr1time gr2time

1	1	1	1	2	29	9	27	65	1	0	65	0
2	1	2	1	2	22	12	48	298	1	0	298	0
3	1	3	1	2	33	6	9	545	1	0	545	0
4	1	4	1	2	11	32	8	785	1	0	785	0
5	2	1	1	2	31	13	91	90	1	0	90	0

For this class, the procedure we will use is `proc mixed`. This procedure fits models with both fixed and random effects. For our model, the SAS code is

```
proc mixed data=back method=ml dfbw;
     class id j;
     model altvas = gr1 t gr1time / s cl;
     repeated / subject=id type=sp(pow)(t) r;
     contrast 'F test for group'
        gr1 1,
        gr1time 1;
     run;
```

Output:

The Mixed Procedure

Model Information

Data Set	WORK.BACK
Dependent Variable	altvas
Covariance Structure	Spatial Power
Subject Effect	id
Estimation Method	ML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Class Level Information

Class	Levels	Values
id	27	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
j	4	1 2 3 4

The Mixed Procedure

Dimensions

Covariance Parameters	2
Columns in X	4
Columns in Z	0

Subjects	27
Max Obs Per Subject	4
Observations Used	105
Observations Not Used	3
Total Observations	108

Iteration History

Iteration	Evaluations	-2 Log Like	Criterion
0	1	965.33552261	
1	2	965.33536606	0.00000072
2	1	965.32661591	0.00248637
3	1	964.87545796	9.36081278
4	1	960.82458648	1054.3531511

The Mixed Procedure

Iteration History

Iteration	Evaluations	-2 Log Like	Criterion
5	1	951.76501329	0.67958481
6	1	948.14411922	0.00055454
7	1	947.91441989	0.00001533
8	1	947.90851534	0.00000002
9	1	947.90850966	0.00000000

Convergence criteria met.

Estimated R Matrix for id 1

Row	Col1	Col2	Col3	Col4
1	579.66	264.96	115.55	51.5904
2	264.96	579.66	252.79	112.86
3	115.55	252.79	579.66	258.80
4	51.5904	112.86	258.80	579.66

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
SP(POW)	id	0.9966
Residual		579.66

Fit Statistics

-2 Log Likelihood	947.9
AIC (smaller is better)	959.9
AICC (smaller is better)	960.8
BIC (smaller is better)	967.7

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	17.43	<.0001

The Mixed Procedure

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	80.5824	7.3880	25	10.91	<.0001	0.05	65.3666	95.7982
gr1	-26.6989	10.2517	25	-2.60	0.0153	0.05	-47.8126	-5.5851
t	-0.03297	0.01277	76	-2.58	0.0117	0.05	-0.05841	-0.00754
gr1time	0.01371	0.01769	76	0.77	0.4409	0.05	-0.02153	0.04895

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gr1	1	25	6.78	0.0153
t	1	76	6.67	0.0117
gr1time	1	76	0.60	0.4409

The Mixed Procedure

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
F test for group	2	76	5.43	0.0062