Note Set #1 for Thursday March 5, 1998

Log-Linear Models in SAS

There are four SAS procedures that can be used almost immediately to fit and evaluate log-linear models. They are

Proc Name	Descripion
PROC FREQ	produces tables similar to `crosstabs()` in SPLUS
PROC CATMOD	fits a variety of models (including log-linear) to categorical data
PROC GENMOD	essentially equivalent to `glm()` in SPLUS
PROC INSIGHT	does a variety of interactive graphics and model fitting.

To illustrate these different procedures I will do partial analyses of the automobile accident data from last time in SAS.

Running an analysis in SAS

To run a SAS job, create the command file, let's say you call it `bozo.sas`, in a text editor. Now there are three things you can do:

1. Say `sas bozo.sas` at the UNIX command prompt, and then after the command finishes running, look for newly created files in your directory, like `bozo.log` and `bozo.lst`.

   Or...

2. In the interactive windows version of SAS, find the small window labelled SAS: PROGRAM EDITOR–Untitled and use the File/Open... pulldown menu to read in your progam to the editor. Once it has been read in, find the small horizontal toolbar labelled Toolbox: PROGRAM EDITOR again and click on the running man in the middle of the toolbar to run the job. The "bozo.log" file will appear in the SAS: LOG window, and the "bozo.lst" file will appear in the SAS: OUTPUT window.

   Or...

3. Even more interesting is to use PROC INSIGHT to do an entire analysis. You can get help on PROC INSIGHT by selecting <u>MODELING AND ANALYSIS TOOLS</u> and then <u>Interactive Data Analysis</u>.

To quit the interactive version of SAS, find the SAS: Session Management window (usually minimized, so you'll have to restore it from its screen icon) and click on the "Terminate" button.

I have combined all the illustrative analyses in a single file here; however you might want to consider doing each one separately. Sample files for the separate analyses can be found in the files `freq.sas`, `catmod.sas`, `genmod.sas` and `insight.sas`, in the class data directory.

```
/* bozo.sas */

data accident;
input Cell Count Type $ Severity $ Eject $ Weight $ ;
datalines;
 1  350 Type1 Sev1 Ejct1  Wght1
 2  150 Type2 Sev1 Ejct1  Wght1
 3   60 Type1 Sev2 Ejct1  Wght1
 4  112 Type2 Sev2 Ejct1  Wght1
 5   26 Type1 Sev1 Ejct2  Wght1
 6   23 Type2 Sev1 Ejct2  Wght1
 7   19 Type1 Sev2 Ejct2  Wght1
 8   80 Type2 Sev2 Ejct2  Wght1
 9 1878 Type1 Sev1 Ejct1  Wght2
10 1022 Type2 Sev1 Ejct1  Wght2
11  148 Type1 Sev2 Ejct1  Wght2
12  404 Type2 Sev2 Ejct1  Wght2
13  111 Type1 Sev1 Ejct2  Wght2
14  161 Type2 Sev1 Ejct2  Wght2
15   22 Type1 Sev2 Ejct2  Wght2
16  265 Type2 Sev2 Ejct2  Wght2
;

proc freq data=accident;
  weight Count;
  tables Type*Severity*Eject*Weight / chisq;
  /* this lets us check conditional independence of Eject and Weight */
  /* given fixed values of Type and Severity (see output) */
run;

proc catmod data=accident;
  weight Count;
  model Type*Severity*Eject*Weight = _response_ /
    ml noresponse noparm pred=freq;
  loglin Type|Severity|Eject|Weight @ 2;
  /* this is like (Type+Severity+Eject+Weight)^2 in SPLUS */
run;

proc genmod data=accident;
   class Type Severity Eject Weight;
   model Count = Type|Severity|Eject|Weight @ 2 /       dist = poisson
                                                        link = log
   /* specify analysis-of-deviance table like SPLUS */  type1;
run;

proc insight data=accident;
run;
```

PROC FREQ

Here is an extract of the output from PROC FREQ:

```
                              The SAS System                          25
                                          11:29 Wednesday, March 4, 1998

                        TABLE 1 OF EJECT BY WEIGHT
                 CONTROLLING FOR TYPE=Type1 SEVERITY=Sev1

                  EJECT     WEIGHT

                  Frequency|
                  Percent  |
                  Row Pct  |
                  Col Pct  |Wght1   |Wght2   |  Total
                  ---------+--------+--------+
                  Ejct1    |    350 |   1878 |   2228
                           |  14.80 |  79.41 |  94.21
                           |  15.71 |  84.29 |
                           |  93.09 |  94.42 |
                  ---------+--------+--------+
                  Total         376     1989     2365
                              15.90    84.10   100.00

                  EJECT     WEIGHT

                  Frequency|
                  Percent  |
                  Row Pct  |
                  Col Pct  |Wght1   |Wght2   |  Total
                  ---------+--------+--------+
                  Ejct2    |     26 |    111 |    137
                           |   1.10 |   4.69 |   5.79
                           |  18.98 |  81.02 |
                           |   6.91 |   5.58 |
                  ---------+--------+--------+
                  Total         376     1989     2365
                              15.90    84.10   100.00

         Statistic                     DF     Value        Prob
         ------------------------------------------------------
         Chi-Square                     1     1.031       0.310

[ much more output omitted ... ]
```

PROC CATMOD

Here is a part of the output from PROC CATMOD:

MAXIMUM-LIKELIHOOD ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
TYPE	1	93.33	0.0000
SEVERITY	1	155.36	0.0000
TYPE*SEVERITY	1	391.60	0.0000
EJECT	1	764.26	0.0000
TYPE*EJECT	1	108.57	0.0000
SEVERITY*EJECT	1	220.36	0.0000
WEIGHT	1	706.50	0.0000
TYPE*WEIGHT	1	15.33	0.0001
SEVERITY*WEIGHT	1	59.17	0.0000
EJECT*WEIGHT	1	1.72	0.1901
LIKELIHOOD RATIO	5	7.33	0.1974

The Chi-squared values here are approximately the change in deviance for adding the corresponding term to the model *last*[1], i.e. this tells you how significant each term remains after accounting for variability due to all other terms in the model.

PROC GENMOD

With the "type1" option on the MODEL line, GENMOD produces exactly the same analysis of deviance table as `glm()` in SPLUS.

LR Statistics For Type 1 Analysis

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	7686.3582	0	.	.
TYPE	7653.6969	1	32.6613	0.0001
SEVERITY	6164.2646	1	1489.4323	0.0001
TYPE*SEVERITY	5562.8450	1	601.4196	0.0001
EJECT	2888.1315	1	2674.7135	0.0001
TYPE*EJECT	2602.1353	1	285.9962	0.0001
SEVERITY*EJECT	2372.8023	1	229.3331	0.0001
WEIGHT	76.3569	1	2296.4453	0.0001
TYPE*WEIGHT	75.5996	1	0.7573	0.3842
SEVERITY*WEIGHT	9.0200	1	66.5797	0.0001
EJECT*WEIGHT	7.3281	1	1.6918	0.1934

[1]You can get exactly these deviances by requesting a "type3" analysis in PROC GENMOD.

PROC INSIGHT

To set this up you need a `bozo.sas` file like this:

```
data accident;
input Cell Count Severity $ Eject $ Weight $ ;
datalines;
 1  350 Type1 Sev1 Ejct1  Wght1
 2  150 Type2 Sev1 Ejct1  Wght1
 3   60 Type1 Sev2 Ejct1  Wght1
 4  112 Type2 Sev2 Ejct1  Wght1
 5   26 Type1 Sev1 Ejct2  Wght1
 6   23 Type2 Sev1 Ejct2  Wght1
 7   19 Type1 Sev2 Ejct2  Wght1
 8   80 Type2 Sev2 Ejct2  Wght1
 9 1878 Type1 Sev1 Ejct1  Wght2
10 1022 Type2 Sev1 Ejct1  Wght2
11  148 Type1 Sev2 Ejct1  Wght2
12  404 Type2 Sev2 Ejct1  Wght2
13  111 Type1 Sev1 Ejct2  Wght2
14  161 Type2 Sev1 Ejct2  Wght2
15   22 Type1 Sev2 Ejct2  Wght2
16  265 Type2 Sev2 Ejct2  Wght2
;
proc insight data=accident;
run;
```

When you run this in interactive SAS, a new window will pop up with the data in a spreadsheet format.

- You can use the Edit/Variables... pulldown menu to recode variables if you wish, or use the Analyze pulldown menu to make a variety of exploratory plots (e.g. scatter plots, box plots, and so forth.
- More interestingly, you can use the Analyze/Fit(Y X)... pulldown menu to perform linear regressions, log-linear model fits, etc. (after selecting the Y and X variables in the dialog box for Analyze/Fit(Y X)..., be sure to click on the Method button and select Poisson for the response distribution and log for the link function, from the menu of radio buttons that appears.)

Getting help in SAS

After starting the interactive version of SAS by typing "sas" at the UNIX prompt on Andrew, find the small horizontal toolbar labelled Toolbox: PROGRAM EDITOR–Untitled and click the help button on the far right.

In the help window that pops up, you can

- Click MODELING AND ANALYSIS TOOLS, then Data Analysis to get a list of PROC's that you can get help on, including CATMOD, GENMOD and FREQ.
- Click SAMPLE PROGRAMS, then SAS/STAT, then e.g. Documentation Examples from Proc Catmod (examples 4 and 5 illustrate fitting log-linear models) or Documentation Examples from Proc Genmod: 1 (look for the Poisson regression example), etc., to get examples of how to set up SAS procedure files to use these commands.
- Graphics aren't all that easy to figure out in SAS. Probably the best thing to do is to try PROC INSIGHT. There are a couple of nice examples of using PROC INSIGHT for EDA and residual analysis in the 36-462/745 web pages this semester:
  - `http://lib.stat.cmu.edu/~giovanni/745/sas_eda/sas_eda.html`
  - `http://lib.stat.cmu.edu/~giovanni/745/sas_reg/sas_reg.html`