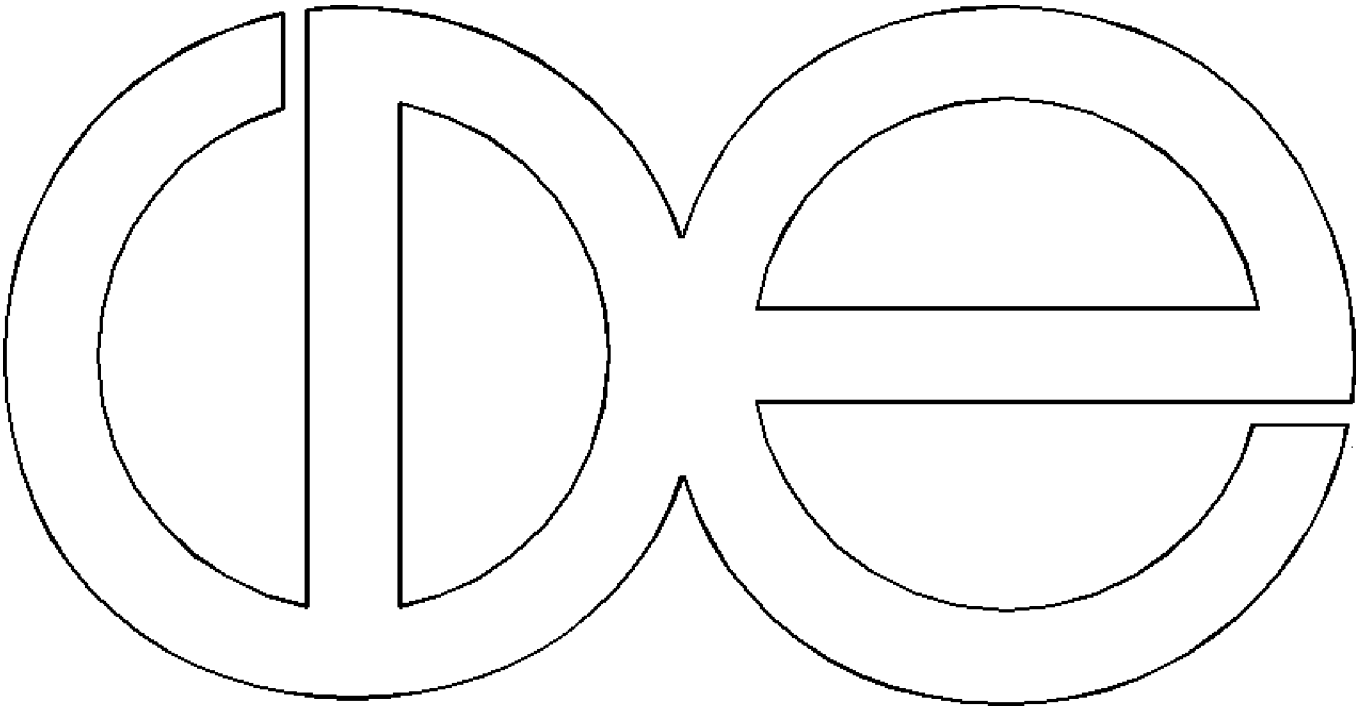
Center for Demography and Ecology

University of Wisconsin-Madison

A Comparative Evaluation of Selected Statistical Software for Computing Multinomial Models

Nancy McDermott

CDE Working Paper No. 95-01



A Comparative Evaluation of Selected Statistical Software for Computing Multinomial Models

Nancy McDermott

Social Science Computing Cooperative
University of Wisconsin - Madison

January 1995

The Center for Demography & Ecology receives core support for Population Research from the Institute for Child Health and Human Development (P30 HD05876).

Table of Contents

1. Introduction

This paper is a comparative evaluation of statistical software for computing multinomial models. The following statistical packages were included in the evaluation: STATA (version 3.1), SAS (version 6.09), and LIMDEP (version 6.0). SPSS (version 4.0), GLIM (version 3.77), and S-PLUS (version 3.1) could not be included because they do not offer a multinomial procedure. Although not considered in this paper, for multinomial models that have an equivalent loglinear model, GLIM or SPSS's LOGLINEAR procedure could be used to fit these models.

The results from a multinomial analysis are first presented for each of the three software packages. Important and unique features of the analyses are noted. In general, the output is not interpreted, except where clarification is needed or where a statistic provided may not be one commonly used. Following the output, comparisons of package features are provided.

Next, a larger data set is used to compare performance (on UNIX) among the three software packages. Finally, recommendations of the appropriate package to use in certain situations are provided.

The results for each software package are presented in the following order: STATA, SAS, and LIMDEP. The ordering, although not random, is not meant to reflect any ranking of preference. A major motivation for this paper was to compare STATA's performance with other statistical packages used at the Social Science Computing Cooperative. Thus, STATA's output is presented first.

1.1 Data

The first data set analyzed was taken from Fienberg (1977, p. 112) and was derived from a survey of 2400 young men rejected for military service because of failure to pass the Armed Forces Qualification Test (AFQT). This information was used to construct a 3×4×2×2 contingency table. The four variables are SED, the respondent's education (1=grammer school, 2=some high school, and 3=high school graduate); FED, the father's education (1=grammer school, 2=some high school, 3=high school graduate, and 4=unknown educational attainment); AGE, which represents the age of the respondent (0=less than 22 years old and 1=greater than or equal to 22 years old); and BLACK which represents the race of the respondent (0=white and 1=black). For this analysis, this contingency table will be used for a multinomial model where the respondent's education (SED) represents the response variable and the remaining variables represent the explanatory variables. The data are summarized below:

Race	Age	Father's Education	Respondent's Education No HS	Some HS	HS Grad
White	< 22	1	39	29	8
		2	4	8	1
		3	11	9	6
		4	48	17	8
	$ 22	1	231	115	51
		2	17	21	13
		3	18	28	45
		4	197	111	35
Black	< 22	1	19	40	19
		2	5	17	5
		3	2	14	3
		4	49	79	24
	$ 22	1	110	133	103
		2	18	38	25
		3	11	25	18
		4	178	206	81

One of the models Fienberg considered was the model with BLACK, AGE, FED, and the interaction between BLACK and FED. This is the model that was fit for each of the three packages. For simplicity, this model will be referred to using a notation that lists only the highest order effects for each variable. Using this notation and abbreviating each effect to its first letter, the above model will be referred to as (A,BF) for the rest of this discussion.

The second example used for performance comparisons is an extract based on the 5% PUMS. The variables include five occupation/industry categories, age in years, educational attainment in years, sex with two categories, race with two categories, and time with two categories (1980 or 1990). The variable representing the five occupation/industry categories was used as the dependent variable. There were 28,369 observations in the extract.

2. Multinomial Models in STATA

The MLOGIT command can be used to fit a multinomial model in STATA. When the response variable takes on more than two outcomes and the outcomes are ordered, then OLOGIT is the appropriate command. This paper concentrates on models where the outcomes have no natural ordering so only the MLOGIT command is described below. The maximum number of explanatory variables that can be fit in any of STATA's estimation procedures is 400.

The following command was used to read in the AFQT data:

```
infile wt sed fed ge22 black using ~/stata/afqt48.dat
```

The MLOGIT command does not generate the indicator variable corresponding to the explanatory variables automatically so the TABULATE command was used to generate the indicator variables for FED:

```
quietly tabulate fed, gen(fedd)
```

Four indicator variables, FEDD1, FEDD2, FEDD3, and FEDD4, are generated as a result of the above command. Recall that GE22 and BLACK were originally coded 0,1 so no further recoding was necessary. The GENERATE command was used to compute the interaction variables between BLACK and FED. GENERATE was used because TABULATE cannot generate indicator variables for interaction effects.

```
generate bfedd1=black*fedd1
generate bfedd2=black*fedd2
generate bfedd3=black*fedd3
```

The indicator variable corresponding to the interaction between race and the fourth category of father's education was not generated so as not to make the model overdetermined. Following is the MLOGIT command to fit the model (A,BF) and the output:

```
mlogit sed black ge22 fedd1 fedd2 fedd3 bfedd1 bfedd2 bfedd3 [fweight=wt]

Iteration 0:  Log Likelihood =-2410.4013
Iteration 1:  Log Likelihood = -2295.239
Iteration 2:  Log Likelihood =-2292.0638
Iteration 3:  Log Likelihood =  -2292.06
Iteration 4:  Log Likelihood =  -2292.06


Multinomial regression                                Number of obs =   2294
                                                      chi2(16)      = 236.68
                                                      Prob > chi2   = 0.0000
Log Likelihood =   -2292.06                           Pseudo R2     = 0.0491

------------------------------------------------------------------------------
     sed |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
2        |
   black |   .8603512   .1410667      6.099   0.000       .5838656    1.136837
```

```
    ge22 |  -.2389912   .1178992     -2.027   0.043      -.4700694    -.007913
   fedd1 |   .0241554   .1502828      0.161   0.872      -.2703934    .3187042
   fedd2 |   .9657354   .3068804      3.147   0.002        .364261     1.56721
   fedd3 |   .8786654   .2712258      3.240   0.001       .3470726    1.410258
  bfedd1 |   .0557878   .2099857      0.266   0.790      -.3557766    .4673522
  bfedd2 |   -.326256   .4048072     -0.806   0.420      -1.119663    .4671516
  bfedd3 |  -.0111627   .4291892     -0.026   0.979       -.852358    .8300327
   _cons |  -.4543555   .1451268     -3.131   0.002      -.7387988   -.1699121
---------+--------------------------------------------------------------------
3        |
   black |   .9809932   .2034249      4.822   0.000       .5822877    1.379699
    ge22 |   .1977878   .1570041      1.260   0.208      -.1099346    .5055103
   fedd1 |   .2166487   .2191206      0.989   0.323      -.2128197    .6461172
   fedd2 |   1.339105   .3827448      3.499   0.000       .5889389    2.089271
   fedd3 |   2.314872   .2856219      8.105   0.000       1.755063    2.874681
  bfedd1 |   .4883564   .2791805      1.749   0.080      -.0588273     1.03554
  bfedd2 |   -.234307   .4849768     -0.483   0.629      -1.184844    .7162301
  bfedd3 |   -1.06167   .4691513     -2.263   0.024       -1.98119   -.1421507
   _cons |  -1.907404   .2128599     -8.961   0.000      -2.324601   -1.490206
------------------------------------------------------------------------------
(Outcome sed==1 is the comparison group)
```

STATA first prints a brief iteration history for the log likelihood. The loglikelihood at iteration 0 is the log likelihood corresponding to a model where only the two constant terms are fit. The output following the iteration history which is labelled `chi2(16)` is the overall chi-squared statistic which evaluates the null hypothesis that all coefficients in the model, except the constant, equal zero. The equation is

$X^2 = -2 (\ln L(i) - \ln L(f))$

where ln L(i) is the initial log likelihood (log likelihood at iteration 0) and lnL(f) is the log likelihood for the final iteration. `(16)` is the degrees of freedom associated with the test and the output labelled `Prob > chi2 = 0.0000` is the probability associated with this chi-squared test.

The output labelled `Pseudo R2` is 1- ln L(f)/ln L(i). However, pseudo R^2 lacks the straightforward explained-variance interpretation of true R^2 in ordinary least squares regression.

The parameter estimates table is presented next. STATA omitted the first level of the response variable; otherwise the model would be overdetermined. This means that the first level of the response variable (respondent had only grammer school education) is used as the comparison group. Coefficients corresponding to the other two groups of SED measure the relative change to the first group. If you prefer to assign a different group as the comparison group, STATA provides an option BASECATEGORY() for this. No matter which category you choose for the comparison group though, you will get a solution to the same underlying model along with the same predicted probabilities. Just the coefficients will differ because they have different interpretations.

You can also display the estimated coefficients transformed to relative risk ratios (e rather than b). STATA was the only package to provide this output. The option is RRR:

```
mlogit, rrr
```

```
Multinomial regression                                  Number of obs =   2294
                                                        chi2(16)      = 236.68
                                                        Prob > chi2   = 0.0000
Log Likelihood =   -2292.06                             Pseudo R2     = 0.0491
------------------------------------------------------------------------------
         sed |        RRR   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
2            |
       black |   2.363991   .3334803      6.099   0.000       1.792956    3.116893
        ge22 |   .7874218   .0928364     -2.027   0.043       .6249589    .9921183
       fedd1 |   1.024449   .1539571      0.161   0.872       .7630792    1.375344
       fedd2 |   2.626719   .8060884      3.147   0.002        1.43945    4.793256
       fedd3 |   2.407684   .6530261      3.240   0.001       1.414919    4.097013
      bfedd1 |   1.057373   .2220333      0.266   0.790       .7006291    1.595763
      bfedd2 |   .7216205   .2921171     -0.806   0.420       .3263896    1.595443
      bfedd3 |   .9888994   .4244249     -0.026   0.979       .4264083    2.293394
-------------+----------------------------------------------------------------
3            |
       black |   2.667104   .5425554      4.822   0.000       1.790129    3.973704
        ge22 |   1.218704   .1913415      1.260   0.208       .8958927    1.657831
       fedd1 |   1.241908   .2721275      0.989   0.323       .8083018    1.908117
       fedd2 |   3.815626   1.460411      3.499   0.000       1.802075    8.079022
       fedd3 |   10.12363    2.89153      8.105   0.000       5.783813    17.71976
      bfedd1 |   1.629636   .4549624      1.749   0.080       .9428696    2.816627
      bfedd2 |   .7911189   .3836743     -0.483   0.629       .3057938    2.046703
      bfedd3 |   .3458776   .1622689     -2.263   0.024        .137905    .8674905
------------------------------------------------------------------------------
(Outcome sed==1 is the comparison group)
```

For either table, STATA always prints the standard error of the estimate along with a z test and the corresponding 95% confidence interval. If you prefer confidence intervals other than 95%, specify this with the LEVEL() option.

3. Multinomial Models in SAS

The CATMOD (CATegorical data MODeling) procedure can be used to fit multinomial models in SAS. CATMOD fits linear models to functions of response frequencies and uses either maximum-likelihood estimation or weighted least squares estimation.

The following data step commands were used to read in the AFQT data:

```
data one;
   infile "afqt48.dat";
   input wt sed fed ge22 black;
```

Unlike STATA and LIMDEP, the CATMOD procedure in SAS uses the last level of the response variable as the comparison group rather than the first level. So, in order to make the parameter estimates comparable, the following statements were added to the above data step to recode the response variable, SED, so the first and last levels are reversed:

```
   if sed=1 then sed2=3;
      else if sed=2 then sed2=2;
         else if sed=3 then sed2=1;
```

The following SAS statements fits the model (A,BF):

```
proc catmod;
   weight wt;
   model sed2=black ge22 fed black*fed/ ml nogls;
run;
```

CATMOD generates the design matrix for categorical explanatory variables automatically. SAS was the only software package examined that had this feature. Explanatory variables are assumed to be categorical unless declared otherwise with a DIRECT statement. The ML NOGLS options instruct SAS to compute maximum-likelihood estimates instead of weighted-least-squares estimates. The output follows:

CATMOD PROCEDURE

Response: SED2	Response Levels (R)= 3
Weight Variable: WT	Populations (S)= 16
Data Set: ONE	Total Frequency (N)= 2294
Frequency Missing: 0	Observations (Obs)= 48

POPULATION PROFILES

Sample	BLACK	GE22	FED	Sample Size
1	0	0	1	76
2	0	0	2	13
3	0	0	3	26
4	0	0	4	73
5	0	1	1	397
6	0	1	2	51
7	0	1	3	91
8	0	1	4	343
9	1	0	1	78
10	1	0	2	29

11	1	0	3	19
12	1	0	4	152
13	1	1	1	346
14	1	1	2	81
15	1	1	3	54
16	1	1	4	465

RESPONSE PROFILES

Response	SED2
1	1
2	2
3	3

MAXIMUM-LIKELIHOOD ANALYSIS

Iteration	Sub Iteration	-2 Log Likelihood	Convergence Criterion	Parameter Estimates 1	2	3
0	0	5040.4332	1.0000	0	0	0
1	0	4601.5268	0.0871	-0.3476	0.2672	-0.3129
2	0	4584.2301	0.003759	-0.4391	0.2904	-0.3828
3	0	4584.1199	0.000024	-0.4512	0.2883	-0.3895
4	0	4584.1199	2.462E-9	-0.4513	0.2883	-0.3895

	Parameter Estimates					
Iteration	4	5	6	7	8	9
0	0	0	0	0	0	0
1	-0.4426	-0.0335	0.1458	-0.3683	-0.4156	0.2884
2	-0.3914	-0.0939	0.1185	-0.4045	-0.3773	0.3468
3	-0.3950	-0.0988	0.1195	-0.4059	-0.3799	0.3552
4	-0.3950	-0.0989	0.1195	-0.4059	-0.3799	0.3552

	Parameter Estimates					
Iteration	10	11	12	13	14	15
0	0	0	0	0	0	0
1	0.4344	0.7767	0.4450	-0.3322	-0.1092	0.0241
2	0.3674	0.9092	0.4378	-0.3417	-0.0612	0.008943
3	0.3707	0.9173	0.4412	-0.3452	-0.0631	0.0161
4	0.3707	0.9173	0.4411	-0.3451	-0.0631	0.0162

	Parameter Estimates		
Iteration	16	17	18
0	0	0	0
1	0.2026	0.4293	0.0165
2	0.1238	0.4219	-0.0325
3	0.1279	0.4298	-0.0296
4	0.1279	0.4299	-0.0296

MAXIMUM-LIKELIHOOD ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
INTERCEPT	2	71.45	0.0000
BLACK	2	32.38	0.0000

GE22	2	9.64	0.0081
FED	6	76.92	0.0000
BLACK*FED	6	18.14	0.0059
LIKELIHOOD RATIO	14	18.10	0.2024

ANALYSIS OF MAXIMUM-LIKELIHOOD ESTIMATES

Effect	Parameter	Estimate	Standard Error	Chi-Square	Prob
INTERCEPT	1	-0.4513	0.0974	21.48	0.0000
	2	0.2883	0.0811	12.62	0.0004
BLACK	3	-0.3895	0.0840	21.52	0.0000
	4	-0.3950	0.0742	28.30	0.0000
GE22	5	-0.0989	0.0785	1.59	0.2077
	6	0.1195	0.0590	4.11	0.0427
FED	7	-0.4059	0.1080	14.14	0.0002
	8	-0.3799	0.0925	16.87	0.0000
	9	0.3552	0.1769	4.03	0.0446
	10	0.3707	0.1533	5.84	0.0156
	11	0.9173	0.1715	28.61	0.0000
	12	0.4411	0.1614	7.47	0.0063
BLACK*FED	13	-0.3451	0.1078	10.25	0.0014
	14	-0.0631	0.0924	0.47	0.4946
	15	0.0162	0.1769	0.01	0.9270
	16	0.1279	0.1533	0.70	0.4041
	17	0.4299	0.1714	6.29	0.0122
	18	-0.0296	0.1614	0.03	0.8544

CATMOD first prints some summary information you can use to verify that you have the right number of observations and that you have specified the model correctly. The table labelled POPULATION PROFILES corresponds to the three-way table in a list format for a crosstabulation of the three explanatory variables, BLACK, GE22, and FED. The table labelled RESPONSE PROFILES lists the levels of the response variable, SED2. Recall that SED2 is simply a reordering of the SED variable.

Next comes the iteration history for the maximum likelihood analysis. The value for -2 Log Likelihood for the final iteration, 4584.1199, when divided by two corresponds to loglikelihood reported by STATA and LIMDEP.

The MAXIMUM-LIKELIHOOD ANALYSIS-OF-VARIANCE TABLE contains a chi-squared test for each effect. The output labelled LIKELIHOOD RATIO compares the specified model with the saturated model and is an appropriate goodness-of-fit test for the model.

Notice that the parameter estimates are different from those of STATA. This is because CATMOD uses a different parameterization for the explanatory variables than was used for STATA. CATMOD constrains the parameters to sum to zero. In other words, CATMOD uses a full-rank center-point parameterization to build design matrices. For example, when the race variable, BLACK, is specified as a categorical variable, each value gets coded internally as either 1 or -1 instead of 1 or 0, as was done for the STATA output in the last section. The sum-to-zero contraint requires that the last level of an effect be the negative of the sum of the other levels of the effect.

If you do not want the full-rank center-point parameterization that CATMOD uses, you can construct the indicator variables yourself in the data step and then insert a DIRECT statement which instructs CATMOD to treat the variables specified as quantitative rather than qualitative. This is illustrated below:

```
data one;
   infile "afqt48.dat";
   input wt sed fed ge22 black;

/* Recode response variable so first level becomes the reference level */
   if sed=1 then sed2=3;
   else if sed=2 then sed2=2;
   else if sed=3 then sed2=1;

/* Generate indicator variables for FED /*
   if fed=1 then fed1=1; else fed1=0;
   if fed=2 then fed2=1; else fed2=0;
   if fed=3 then fed3=1; else fed3=0;

/* Generate indicator variables for FED*BLACK */
   bfed1=fed1*black;
   bfed2=fed2*black;
   bfed3=fed3*black;
run;

proc catmod;
   direct black ge22 fed1 fed2 fed3 bfed1 bfed2 bfed3;
   weight wt;
   model sed2=black ge22 fed1 fed2 fed3 bfed1 bfed2 bfed3 / ml nogls;
run;
```

If you execute the above code, the parameter estimates will be identical to those shown earlier for STATA. No matter which way you choose to specify the model, you will get a solution to the same underlying model along with the same predicted probabilities.

4. Multinomial Models in LIMDEP

The LOGIT command in LIMDEP fits both logit models and multinomial models. The maximum number of explanatory variables that can be fit in LOGIT is 150. As with the MLOGIT command in STATA, LOGIT will not generate indicator variables for you. You must create them manually with the CREATE command. The following commands were used to read in the AFQT data and create the indicator variables for father's education:

```
batch
read; nrec=48; nvar=5; file=afqt48.dat;
      names=wt,sed,fed,ge22,black$
create; if (fed=1) fedd1=1;
        if (fed=2) fedd2=1;
        if (fed=3) fedd3=1;
        bfedd1=fedd1*black;
        bfedd2=fedd2*black;
        bfedd3=fedd3*black $
```

LOGIT requires that the response variable be coded 0, 1, and so on. Since SED was coded 1, 2, or 3, it had to be recoded:

```
create; sed=sed-1$
```

Following is the MLOGIT command to fit the model (A,BF) and the output:

```
logit; lhs=sed; rhs=one,black,ge22,fedd1,fedd2,fedd3,bfedd1,bfedd2,bfedd3;
       wts=wt,noscale$

 Multinomial Logit Model
 3 Outcomes: SED=0    SED=1    SED=2
 Coefficients for SED=0    set to zero.
 Least squares starting values:
 Dep. Var. is binary: SED=1
 N(0,1) used for significance levels.

 Variable  Coefficient  Std. Error   t-ratio Prob:t:>x   Mean of X  St.Dv.of X
--------------------------------------------------------------------------------
Constant  0.36727       0.3138E-01   11.703  0.00000
BLACK     0.14910       0.3059E-01    4.874  0.00000  0.53357       0.49898
GE22     -0.72252E-01   0.2509E-01   -2.880  0.00398  0.79686       0.40242
FEDD1    -0.21829E-02   0.3236E-01   -0.067  0.94621  0.39102       0.48809
FEDD2     0.14344       0.6464E-01    2.219  0.02648  0.75850E-01   0.26482
FEDD3     0.51698E-02   0.5039E-01    0.103  0.91828  0.82825E-01   0.27568
BFEDD1   -0.47203E-01   0.4439E-01   -1.063  0.28759  0.18483       0.38824
BFEDD2   -0.10660       0.8160E-01   -1.306  0.19147  0.47951E-01   0.21371
BFEDD3    0.66158E-01   0.7802E-01    0.848  0.39647  0.31822E-01   0.17556

 Multinomial Logit Model
 3 Outcomes: SED=0    SED=1    SED=2
 Coefficients for SED=0    set to zero.
 Least squares starting values:
 Dep. Var. is binary: SED=2
 N(0,1) used for significance levels.

 Variable  Coefficient  Std. Error   t-ratio Prob:t:>x   Mean of X  St.Dv.of X
--------------------------------------------------------------------------------
Constant  0.64364E-01   0.2518E-01    2.557  0.01057
BLACK     0.70165E-01   0.2454E-01    2.859  0.00425  0.53357       0.49898
```

```
GE22      0.47302E-01   0.2013E-01    2.350  0.01876  0.79686       0.40242
FEDD1     0.20670E-01   0.2596E-01    0.796  0.42587  0.39102       0.48809
FEDD2     0.11669       0.5185E-01    2.250  0.02443  0.75850E-01   0.26482
FEDD3     0.33474       0.4042E-01    8.281  0.00000  0.82825E-01   0.27568
BFEDD1    0.93936E-01   0.3561E-01    2.638  0.00834  0.18483       0.38824
BFEDD2    0.48561E-02   0.6547E-01    0.074  0.94087  0.47951E-01   0.21371
BFEDD3   -0.21659       0.6259E-01   -3.460  0.00054  0.31822E-01   0.17556
 Iterations: Method=NEWTON   Maximum iterations  25
Convergence criteria:       Gradient= 0.100D-03    F= 0.100D-03  b= 0.100D-04

********************************************************************************

                        Method=NEWTON; Maximum iterations 25
   Convergence criteria: Gradient=   0.1000000E-03
                         Function  =   0.1000000E-03
                         Parameters=   0.1000000E-04
   Starting values:   0.3673      0.1491     -0.7225E-01 -0.2183E-02  0.1434
                      0.5170E-02 -0.4720E-01 -0.1066      0.6616E-01
0.6436E-01                       0.7017E-01  0.4730E-01  0.2067E-01  0.1167
0.3347                           0.9394E-01  0.4856E-02 -0.2166

 ==> NEWTON ITERATIONS

 Iteration:   1 Fn=   2531.151
 PARAM   0.367     0.149    -0.723E-01-0.218E-02 0.143     0.517E-02-0.472E-01
        -0.107     0.662E-01 0.644E-01 0.702E-01 0.473E-01 0.207E-01 0.117
         0.335     0.939E-01 0.486E-02-0.217
 GRADNT   26.4     -49.9      43.5      33.0     -12.1     -4.19     -6.82
         -9.50     -8.17      309.      122.      239.      120.      11.8
         -1.73      25.3      4.91      3.10

 Iteration:   2 Fn=   2301.726
 PARAM  -0.736      1.10    -0.263     0.510E-01  1.24      1.19     0.751E-01
        -0.545    -0.282     -1.69     0.946     0.871E-01 0.133      1.26
          2.19     0.487    -0.254     -1.07
 GRADNT  -67.9     -14.4     -53.4     -27.9    -0.470     0.156     -1.39
         0.166     0.741E-01  28.4      11.9      18.2      7.75     0.465
        -0.315     0.968    -0.161    -0.356

 Iteration:   3 Fn=   2292.113
 PARAM  -0.429     0.835    -0.238     0.171E-01 0.939     0.852     0.617E-01
        -0.301     0.143E-01 -1.88     0.970     0.190     0.205      1.32
          2.30     0.495    -0.225     -1.05
 GRADNT   3.91    -0.154      3.53      1.89    -0.128E-01-0.121E-01-0.198E-01
        -0.686E-02-0.728E-02 0.596     0.482     0.511E-01-0.130     0.198E-01
         0.841E-02 0.242E-01 0.665E-02 0.519E-02

 Iteration:   4 Fn=   2292.060
 PARAM  -0.454     0.860    -0.239     0.241E-01 0.966     0.879     0.558E-01
        -0.326    -0.111E-01 -1.91     0.981     0.198     0.217      1.34
          2.31     0.488    -0.234     -1.06
 GRADNT  0.142E-01-0.100E-02 0.137E-01 0.671E-02-0.784E-04-0.528E-04-0.122E-03
        -0.531E-04-0.390E-04-0.267E-02 0.155E-02-0.373E-02-0.174E-02 0.127E-03
         0.965E-04 0.240E-03 0.820E-04 0.553E-04

 Iteration:   5 Fn=   2292.060
 PARAM  -0.454     0.860    -0.239     0.242E-01 0.966     0.879     0.558E-01
        -0.326    -0.112E-01 -1.91     0.981     0.198     0.217      1.34
          2.31     0.488    -0.234     -1.06
 GRADNT  0.240E-06-0.571E-08 0.223E-06 0.919E-07-0.644E-09-0.107E-09-0.423E-09
        -0.492E-09-0.423E-09 0.371E-07 0.129E-07 0.249E-07 0.552E-08 0.132E-08
         0.909E-09 0.222E-08 0.913E-09 0.651E-09
 ** Gradient has converged.
 ** Function has converged.
 ** B-vector has converged.
```

```
********************************************************************************
 Maximum Likelihood Estimates
 Log-Likelihood.............     -2292.1
 Restricted (Slopes=0) Log-L.     -2410.4
 Chi-Squared (16)............      236.68
 Significance Level.........     0.32173E-13
 N(0,1) used for significance levels.
 Variable  Coefficient  Std. Error   t-ratio Prob:t:>x   Mean of X  St.Dv.of X
--------------------------------------------------------------------------------
Constant -0.45436       0.1451       -3.131  0.00174
BLACK     0.86035       0.1411        6.099  0.00000  0.53357       0.49898
GE22     -0.23899       0.1179       -2.027  0.04265  0.79686       0.40242
FEDD1     0.24155E-01   0.1503        0.161  0.87230  0.39102       0.48809
FEDD2     0.96574       0.3069        3.147  0.00165  0.75850E-01   0.26482
FEDD3     0.87867       0.2712        3.240  0.00120  0.82825E-01   0.27568
BFEDD1    0.55788E-01   0.2100        0.266  0.79049  0.18483       0.38824
BFEDD2   -0.32626       0.4048       -0.806  0.42027  0.47951E-01   0.21371
BFEDD3   -0.11163E-01   0.4292       -0.026  0.97925  0.31822E-01   0.17556
Constant  -1.9074       0.2129       -8.961  0.00000
BLACK     0.98099       0.2034        4.822  0.00000  0.53357       0.49898
GE22      0.19779       0.1570        1.260  0.20776  0.79686       0.40242
FEDD1     0.21665       0.2191        0.989  0.32280  0.39102       0.48809
FEDD2      1.3391       0.3827        3.499  0.00047  0.75850E-01   0.26482
FEDD3      2.3149       0.2856        8.105  0.00000  0.82825E-01   0.27568
BFEDD1    0.48836       0.2792        1.749  0.08025  0.18483       0.38824
BFEDD2   -0.23431       0.4850       -0.483  0.62900  0.47951E-01   0.21371
BFEDD3    -1.0617       0.4692       -2.263  0.02364  0.31822E-01   0.17556
Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

            Predicted

Actual       0     1     2       TOTAL

  0          4    11     1          16
  1          4    11     1          16
  2          4    11     1          16

Total       12    33     3          48
```

The LOGIT output begins with results from a least squares estimation which LIMDEP uses to obtain starting values for the maximum likelihood estimation. The iteration history for the maximum likelihood estimates appear next. The output labelled `Restricted Log-L` is the log-likelihood function for a model with only a constant. The `Chi-Squared` tests the null hypothesis that the coefficients for all the terms in the model except the constant are zero.

Like STATA, LIMDEP omits the first level of the response variable so as not to let the model be overdetermined. This means that the first level of the response variable (respondent had only grammer school education) is used as the comparison group. Coefficients corresponding to the other two groups of SED measure the relative change to the first group. LIMDEP prints the standard error of the estimate along with a t-test for the null hypothesis that the parameter estimate is zero.

Last, LIMDEP prints a contingency table of the predicted outcomes from the fitted model by the actual outcomes.

5. Feature Comparisons

The table below summarizes the features of the multinomial analyses for each of the software packages examined.

Key: **feature available** **feature not available**	**STATA**	**SAS**	**LIMDEP**
Model Specification			
Automatic Dummy Variable Generation			
Maximum Number of Parameters that can be Fit	400	none	150
Shortcut Syntax for Specifying Complex Models			
Specify Value of Dependent Variable to be Treated as Base Category	T	T	
Option for Computing a Conditional Multinomial Model	T	T	T
Test Hypotheses about Coefficients in Model	T	T	
Allows for alternative sampling schemes			
Intercept Suppression Option			
Maximum Likelihood Estimates			
Relative Risk Ratios	T		
Chi-square Statistic for Parameter Estimate			
t-statistic or z-statistic for Parameter Estimate	T		
Confidence Intervals for Parameter Estimates	T		
Confidence Intervals for Odds Ratios	T		
Regression Diagnostics	**T**	**T**	**T**
Criteria for Assessing Fit			
Loglikelihood	T	T	T
Goodness-of-Fit Test			
Pseudo R-Square	T		
Correlation Matrix of Parameter Estimates	**T**	**T**	**T**

6. Performance Comparisons

The PUMS extract was used for the performance comparisons. These data contained 28,369 observations. The variable representing the five occupation/industry categories was used as the dependent variables. The explanatory variables include age in years, educational attainment in years, sex with two categories, race with two categories, and time with two categories (1980 or 1990). A full five-way model was fit which required that 128 parameters be fit.

The UNIX `time` command was used to compare the performances of the statistical packages. Each program for each package (for both the small and large data set) was run 10 times. The average time in seconds spent in execution of the program (not real time) are shown in the table below. Execution times for the smaller AFQT data set were included to give an indication of a package's "overhead." For example, SAS is a huge program with lots of overhead and the times reported for the AFQT data represent this.

All runs were done on one machine (a Sparc 10/512 MP with 128 Mb memory) because a machine's processing speed would affect the time reported.

	AFQT Data (time in seconds)	PUMS Data (time in minutes)
STATA	0.18 (1)	28.68 (2)
SAS	0.4 (2)	1.97 (1)
LIMDEP	0.82 (3)	218.54 (3)

The number in parentheses represents the package's relative rank for performance.

7. Recommendations

All the statistical packages gave the same results within round-off error. Some packages reported more significant digits than others. However, this does not indicate that the package computed statistics with more accuracy. You should be conservative about the number of significant digits you report from statistical packages; usually the package reports too many digits leading you to believe they are all significant when they are not.

The complexity and number of effects in the model will probably be the deciding factors in your decision about which package to use. For complex models, the multinomial procedures provided by LIMDEP and STATA are not a good choice because they do not construct the indicator variables for the design matrix automatically. For models with a lot of effects, you will also want to avoid LIMDEP and STATA because of their restriction on the number of parameters that can be fit. LIMDEP has a limit of 150 parameters and STATA has a limit of 400.

SAS was also the clear winner when it came to performance. You could still compute a large model in a reasonable amount of time with STATA but you would definitely want to avoid LIMDEP for large models.

8. References

Agresti, A. (1990), Categorical Data Analysis, New York: John Wiley and Sons, Inc.

Fienberg, Stephen E. (1977), The Analysis of Cross-Classified Categorical Data, MIT Press.

Greene, William H. (1992), Limdep User's Manual and Reference Guide: Version 6, Bellport, NY: Econometric Software, Inc.

SAS Institute Inc. (1989), SAS/STAT Users Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute Inc.

Stata Corporation (1993), Stata Reference Manual: Release 3.1, College Station, TX

Center for Demography & Ecology
University of Wisconsin
1180 Observatory Drive, Rm. 4412
Madison WI 53706-1393
U.S.A.
608/262/2182
FAX 608/262/8400