Multiple matrix sampling data

This example illustrates the use of the TYPE=3 specification on the INPUT command to analyze aggregate-level, multiple-matrix sampling data. The data in **exampl06.dat** are numbers tried and numbers correct for items from eight forms of a matrix sampled assessment instrument. The groups are selected 8th grade students from 32 public schools. The first record for each school contains the data for the items of a Number Concepts scale, NUMCON, and the second record contains the data for items of an Algebra Concepts scale, ALGCON. Data for the first two schools are shown below.

```
SCHOOL   1 NUM    1   0   3   2   2   1   4   4   3   2   2   1   4   3   4   1
SCHOOL   1 ALG    1   0   3   1   2   0   3   2   3   2   2   1   4   1   4   0
SCHOOL   2 NUM    5   3   4   4   3   2   3   3   2   2   4   3   4   3   5   3
SCHOOL   2 ALG    5   2   4   2   3   2   3   2   2   2   4   2   4   2   5   3
```

An answer key is not required for aggregate-level data in number-tried, number-right summary form. Note the format statement for reading the two sets of eight number-tried, number-right observations.

The items are multiple-choice and fairly difficult, so the 3PL model is needed. Because aggregate-level data are always more informative than individual-level item responses, it is worthwhile in the CALIB command to increase the number of quadrature points (NQPT), to set a stricter criterion for convergence (CRIT), and to increase the CYCLES limit. A prior on the thresholds (TPRIOR) and a ridge constant of 0.8 (RIDGE) are required for convergence with the exceptionally difficult ALGCON subtest. Aggregate-level data typically have smaller slopes in the 0,1 metric than do person-level data. For this reason, the mean of the prior for the log slopes has been set to 0.5 by use of the READPRIOR option of the CALIB command and the following PRIOR commands.

The aggregate scores for the schools are estimated by the EAP method using the empirical distributions from Phase 2. The number of quadrature points is set the same as in Phase 2.

The scores are rescaled to a mean of 250 and a standard deviation of 50 in the latent distribution of schools (IDIST=3, LOCATION=250, SCALE=50). The fit of the data to the group-level model is tested for each school (FIT). The NUMCON items have fairly homogeneous slopes and might be favorable for a one-parameter model.

```
EXAMPL06.BLM - MULTIPLE-MATRIX SAMPLING DATA
               AGGREGATE-LEVEL MODEL
>GLOBAL   NPARM=3, NTEST=2, DFNAME='EXAMPL06.DAT';
>LENGTH   NITEMS=(8,8);
>INPUT    NTOTAL=16, NALT=5, NIDCHAR=9, TYPE=3;
>ITEMS    INUM=(1(1)16), INAMES=(N1(1)N8,A1(1)A8);
>TEST1    TNAME=NUMCON, INUM=(1(1)8);
>TEST2    TNAME=ALGCON, INUM=(9(1)16);
(9A1,T15,8(2F3.0)/T15,8(2F3.0))
>CALIB    NQPT=51, CYCLES=50, NEWTON=10, CRIT=0.005, TPRIOR,
          READPRIOR, NOFLOAT, RIDGE=(2,0.8,2.0), CHI=8, PLOT=1;
>PRIORS1  SMU=(0.5(0)8);
```

```
>PRIORS2  SMU=(0.5(0)8);
>SCORE    NQPT=(12,12), IDIST=3, RSCTYPE=4,
          LOCATION=(250.0,250.0), SCALE=(50.0,50.0), FIT;
```

Phase 1 output

Group-level data consist of number-tried and number-right frequencies for each item in each group. The program reads them as values rather than characters and conversion to item scores is not required.

```
OBSERVATION #     1  WEIGHT:     1.0000  ID : SCHOOL  1

 SUBTEST #:   1    NUMCON
   GROUP #:   1

     TRIED    RIGHT
    23.000   14.000
 ITEM       1     2     3     4     5     6     7     8
 TRIED    1.0   3.0   2.0   4.0   3.0   2.0   4.0   4.0
 RIGHT    0.0   2.0   1.0   4.0   2.0   1.0   3.0   1.0

 SUBTEST #:   2    ALGCON
   GROUP #:   1

     TRIED    RIGHT
    22.000    7.000
 ITEM       1     2     3     4     5     6     7     8
 TRIED    1.0   3.0   2.0   3.0   3.0   2.0   4.0   4.0
 RIGHT    0.0   1.0   0.0   2.0   2.0   1.0   1.0   0.0

 OBSERVATION #     2  WEIGHT:     1.0000  ID : SCHOOL  2

 SUBTEST #:   1    NUMCON
   GROUP #:   1

     TRIED    RIGHT
    30.000   23.000
 ITEM       1     2     3     4     5     6     7     8
 TRIED    5.0   4.0   3.0   3.0   2.0   4.0   4.0   5.0
 RIGHT    3.0   4.0   2.0   3.0   2.0   3.0   3.0   3.0

 SUBTEST #:   2    ALGCON
   GROUP #:   1

     TRIED    RIGHT
    30.000   17.000
 ITEM       1     2     3     4     5     6     7     8
 TRIED    5.0   4.0   3.0   3.0   2.0   4.0   4.0   5.0
 RIGHT    2.0   2.0   2.0   2.0   2.0   2.0   2.0   3.0
```

Classical item statistics are computed for each subtest. Biserial correlations cannot be computed with group-level data.

```
ITEM STATISTICS FOR SUBTEST NUMCON
ITEM*TEST CORRELATION
ITEM   NAME        #TRIED    #RIGHT   PCT    LOGIT/1.7  PEARSON  BISERIAL
-------------------------------------------------------------------------
    1   N1          260.0     160.0    61.5    -0.28     0.637    0.000
    2   N2          268.0     162.0    60.4    -0.25     0.682    0.000
    3   N3          260.0     163.0    62.7    -0.31     0.663    0.000
    4   N4          261.0     137.0    52.5    -0.06     0.637    0.000
    5   N5          271.0     129.0    47.6     0.06     0.699    0.000
    6   N6          271.0     154.0    56.8    -0.16     0.656    0.000
    7   N7          270.0     157.0    58.1    -0.19     0.656    0.000
```

```
    8   N8          266.0     170.0    63.9    -0.34     0.781    0.000
-------------------------------------------------------------------------

 ITEM STATISTICS FOR SUBTEST ALGCON

                                                        ITEM*TEST CORRELATION
ITEM   NAME        #TRIED    #RIGHT   PCT    LOGIT/1.7  PEARSON  BISERIAL
-------------------------------------------------------------------------
    1   A1          259.0     120.0    46.3     0.09     0.636    0.000
    2   A2          267.0      81.0    30.3     0.49     0.606    0.000
    3   A3          241.0      94.0    39.0     0.26     0.669    0.000
    4   A4          245.0     121.0    49.4     0.01     0.687    0.000
    5   A5          263.0      96.0    36.5     0.33     0.669    0.000
    6   A6          263.0     166.0    63.1    -0.32     0.746    0.000
    7   A7          267.0      71.0    26.6     0.60     0.667    0.000
    8   A8          262.0      90.0    34.4     0.38     0.683    0.000
-------------------------------------------------------------------------
```

Phase 2 output

The set-up for group-level item calibration differs somewhat from examinee-level analysis: more quadrature points and more iterations for the solution are required. Prior distributions for all parameters are necessary, the means should be kept fixed (default = NOFLOAT), and the mean of the priors for slopes should be set lower than the examinee-level default.

```
>PRIORS1 SMU = (0.5000(0)8);

CONSTRAINT DISTRIBUTIONS ON ITEM PARAMETERS
  (THRESHOLDS, NORMAL; SLOPES, LOG-NORMAL; GUESSING, BETA)

                THRESHOLDS          SLOPES          ASYMPTOTES
 ITEM          MU     SIGMA      MU     SIGMA     ALPHA       BETA
 ------------------------------------------------------------------
 N1          0.000    2.000    0.500    1.649       5.00      17.00
 N2          0.000    2.000    0.500    1.649       5.00      17.00
 N3          0.000    2.000    0.500    1.649       5.00      17.00
 N4          0.000    2.000    0.500    1.649       5.00      17.00
 N5          0.000    2.000    0.500    1.649       5.00      17.00
 N6          0.000    2.000    0.500    1.649       5.00      17.00
 N7          0.000    2.000    0.500    1.649       5.00      17.00
 N8          0.000    2.000    0.500    1.649       5.00      17.00
 ------------------------------------------------------------------
```

Group-level item parameter estimates for the first 3 items in subtest NUMCON are as follows.

```
SUBTEST NUMCON  ;  ITEM PARAMETERS AFTER CYCLE   12

ITEM      INTERCEPT    SLOPE    THRESHOLD   LOADING   ASYMPTOTE    CHISQ     DF
             S.E.       S.E.       S.E.       S.E.       S.E.      (PROB)
 ------------------------------------------------------------------------------
 N1       |   0.030  |   0.190  |  -0.156  |   0.186  |   0.232  |     5.7   6.0
          |   0.194* |   0.066* |   1.026* |   0.065* |   0.094* | (0.4521)
          |          |          |          |          |          |
 N2       |   0.046  |   0.279  |  -0.163  |   0.268  |   0.218  |     3.8   6.0
          |   0.222* |   0.107* |   0.801* |   0.103* |   0.093* | (0.7025)
          |          |          |          |          |          |
 N3       |   0.126  |   0.313  |  -0.404  |   0.299  |   0.212  |     3.2   5.0
          |   0.224* |   0.120* |   0.735* |   0.115* |   0.091* | (0.6638)
-------------------------------------------------------------------------------
                                          * STANDARD ERROR

     LARGEST CHANGE =    0.003146                          42.8  53.0
                                                           (0.8397)
```

```
 NOTE: ITEM FIT CHI-SQUARES AND THEIR SUMS MAY BE UNRELIABLE
 FOR TESTS WITH LESS THAN 20 ITEMS

 PARAMETER       MEAN  STN DEV
 ----------------------------------
 ASYMPTOTE      0.210    0.041
 SLOPE          0.306    0.099
 LOG(SLOPE)    -1.223    0.290
 THRESHOLD      2.241    1.515
```

Phase 3 output

Computing scores at the group-level is essentially the same as at the examinee level. Note that the selection of EAP estimations based on the empirical latent distribution from Phase 2 overrides the choice here of number of quadrature points. Because of the small number of items, the standard deviation of the estimated scores is considerably smaller than that of the latent distribution. Portions of the Phase 3 output are listed below.

```
 >SCORE NQPT = (12, 12),IDIST = 3,RSCTYPE = 4,
 LOCATION = (250.0000, 250.0000), SCALE = (50.0000, 50.0000),FIT;

 PARAMETERS FOR SCORING, RESCALING, AND TEST AND ITEM INFORMATION
 METHOD OF SCORING SUBJECTS:                EXPECTATION A POSTERIORI
                                            (EAP; BAYES ESTIMATION)
 TYPE OF PRIOR:                             EMPIRICAL, FROM ITEM CALIBRATION
 SUBJECT FIT PROBABILITIES:                 YES
 TYPE OF RESCALING:                         IN THE ESTIMATED LATENT
                                             DISTRIBUTION
 REFERENCE GROUP FOR RESCALING:             GROUP:  1

                QUAD   RESCALING  CONSTANTS
 TEST    NAME  POINTS    SCALE    LOCATION
 ----------------------------------------
    1  NUMCON     51    50.000   250.000
    2  ALGCON     51    50.000   250.000
-----------------------------------------
```

The scores are rescaled so that the mean and standard deviation of the Phase 3 latent distribution are 250 and 50, respectively. Scores for all 32 schools are computed and printed. Because the data are binomial rather than binary, a χ^2 index of fit on 8 degrees of freedom can be calculated for each school. The corresponding probabilities are shown in the output.

```
 RESCALING WITH RESPECT TO LATENT DISTRIBUTION
 -------------------------------------------------

          RESCALING  CONSTANTS
 TEST       SCALE    LOCATION
 NUMCON      58.462   251.342
 ALGCON      56.462   251.127

GROUP   SUBJECT IDENTIFICATION                              FIT  MARGINAL
WEIGHT  TEST  TRIED  RIGHT  PERCENT     ABILITY      S.E.   PROB   PROB
-------------------------------------------------------------------------
1  SCHOOL  1                        |                      |
1.00   NUMCON    23     14    60.87 |  246.5104   43.5894 | 0.1539 0.0000
1  SCHOOL  1                        |                      |
1.00   ALGCON    22      7    31.82 |  243.1547   47.4683 | 0.3197 0.0000
 [Similar output omitted]
1  SCHOOL 32                        |                      |
1.00   NUMCON   181    100    55.25 |  221.6762   21.9655 | 0.0166 0.0000
1  SCHOOL 32                        |                      |
1.00   ALGCON   179     77    43.02 |  273.1747   21.8821 | 0.5242 0.0000
-------------------------------------------------------------------------
```

MEAN & SD OF SCORE ESTIMATES AFTER RESCALING: 250.000 31.149
MEAN & SD OF LATENT DISTRIBUTION AFTER RESCALING: 250.000 50.000