



## Two-group differential item functioning (DIF) analysis with the partial credit model

This example illustrates differential item functioning (DIF) analysis of multiple category item responses. The SCORE command is required and thus included in the command file. For the DIF model, however, no scoring is done and there is no Phase 3 output.

Raw data are read from the file **exampl04.dat** using the DFNAME keyword on the FILES command. The data file contains responses to 6 items, as indicated on the INPUT command, where NTOTAL is set to 6. The data file contains the examinee ID and sample group code (1,2), then the responses on the 6 items, and finally the generating trait value for each examinee. The first few lines of the data file are shown below.

```
0001 1 233113   .43930
0002 1 113111  -.94251
0003 1 313112  -.57257
0004 1 113131  -.59414
0005 1 313131  -.36019
```

The format statement includes information on three fields in the raw data file. The subject ID (4A1) and group identification field (1A1) are read first, followed by the 6 item responses (6A1).

One test, 6 items in length, is considered. The MGROUP keyword on the INPUT command requests a multiple-group analysis for two groups. Note that the MGROUP keyword is used in combination with the MGROUP command, which must follow directly after the BLOCK command(s).

On the TEST command, a name for the test is provided using the TNAME keyword. The items on this test are listed using the ITEMS keyword, while the INAMES keyword is used to provide names for the items. Finally, by setting NBLOCK to 6, it is indicated that 6 BLOCK commands will follow the TEST command.

In this example, there is one item with three categories originally coded 1, 2, and 3 in each block as indicated by the NITEMS, NCAT and ORIGINAL keywords respectively. Because the rating-scale model is not used here, separate category parameters are estimated for each item, and the REPEAT keyword indicates that the BLOCK command should be repeated six times.

The second value (1) assigned to the DIF keyword of the MGROUP command requests a DIF analysis of the item threshold parameters. All other values in this keyword are equal to zero, indicating that only thresholds are allowed to differ between the groups. The GNAME and GCODE keywords are used to assign names and codes to the two groups. By default, the first group will be used as the reference group. To change the reference group, the REFERENCE keyword on the MGROUP command may be used.

A partial credit model with logistic response function is requested through the use of the PARTIAL and LOGISTIC options on the CALIB command. The default number of quadrature points is 30. In this case, NQPT is set to 25, because fewer points are needed when the number of items is small. By setting the CYCLES keyword to 100, a maximum of 100 EM cycles will be performed, followed by two Newton cycles (NEWTON=2). The convergence criterion is somewhat relaxed by setting CRIT to 0.01 instead of using the default convergence criterion of 0.001. Finally, the POSTERIOR option is added to the CALIB command. As a default, the posterior distribution is computed after the computation of expected proportions during the E-step as their by-product. Therefore, these expected sample sizes and expected frequencies of categorical responses are computed based on the posterior distribution in the previous EM cycle. Adding the POSTERIOR option forces the program to compute the posterior distribution again after the M-step. Therefore, the expected proportions can be computed during the E-step based on an updated posterior distribution. This was added to be consistent with the BILOG-MG program in the case of two categories.

The command file is as follows.

```

EXAMPL04.PSL - DIF ANALYSIS USING PARTIAL CREDIT MODEL
                ARTIFICIAL DATA, TWO SAMPLES (EACH WITH N=500, N=(0,1)), 6 ITEMS
>FILES   DFNAME='EXAMPL04.DAT';
INPUT   NIDCHAR=4, MGROUP=2, NTOTAL=6;
        (4A1,1X,1A1,1X,6A1)
>TEST    TNAME='PARV3E', ITEM=(1,2,3,4,5,6),
        INAME=('I001','I002','I003','I004','I005','I006'), NBLOCK=6;
>BLOCK1 REPEAT=6, NIT=1, NCAT=3, ORIGINAL=(1,2,3);
>MGROUP  DIF=(0,1,0,0), GNAME=('MALE','FEMALE'), GCODE=('1','2');
>CALIB   LOGISTIC, PARTIAL, NQPT=25, CYCLES=(100,1,1,1,1,1), NEWTON=20,
        CRIT=0.01, POSTERIOR;
>SCORE   ;

```

### Phase 0 output

When the MGROUP keyword and MGROUP command are used or multiple TEST/BLOCK commands are used, additional information is written to the phase 0 output file.

```

NUMBER OF SUBGROUPS:    2
FORMAT OF DATA INPUT IS
(4A1,1X,1A1,1X,6A1)

>TEST    TNAME=PARV3E, ITEM=(1,2,3,4,5,6),
        INAME=('I001','I002','I003','I004','I005','I006'), NBLOCK=6 ;

BLOCK CARD:    1
>BLOCK1 REPEAT=6, NIT=1, NCAT=3, ORIGINAL=(1,2,3) ;

```

```

BLOCK CARD:  2 IS COPIED FROM BLOCK  1
BLOCK CARD:  3 IS COPIED FROM BLOCK  1
BLOCK CARD:  4 IS COPIED FROM BLOCK  1
BLOCK CARD:  5 IS COPIED FROM BLOCK  1
BLOCK CARD:  6 IS COPIED FROM BLOCK  1

```

```
>MGROUP DIF=(0,1,0,0), GNAME=(MALE,FEMALE), GCODE=('1','2') ;
```

In the next few lines, the program echoes the information on parameters allowed to be different between groups as specified with the DIF keyword: in this case, only the thresholds are allowed to differ between the two groups. The MALE group will be used as reference group.

```

GROUP PARAMETER FOR SLOPE:          NO
GROUP PARAMETER FOR THRESHOLD:      YES
GROUP PARAMETER FOR CATEGORY:       NO
GROUP PARAMETER FOR GUESSING:       NO
REFERENCE GROUP FOR DIF MODEL:      1  MALE

```

```

SUBGROUP NAME AND CODE
=====

```

```

1  MALE      1
2  FEMALE    2

```

```

DIF OR COMMON BLOCK
=====

```

```

1  BLOCK      DIF BLOCK
2  BLOCK      DIF BLOCK
3  BLOCK      DIF BLOCK
4  BLOCK      DIF BLOCK
5  BLOCK      DIF BLOCK
6  BLOCK      DIF BLOCK

```

### Phase 1 output

The only difference between the Phase 1 output for a single group analysis and for a multiple-group analysis is that the summary item statistics are first given by subgroup and then for the total group. The output for the first item is shown below for all three cases. We see that females were more likely to respond in category 3 and less likely to respond in category 1 than the males. In general, 76% of the total responses were in category 3.

```
1 SUBGROUP: MALE
```

```
BLOCK NO.:  1      NAME: BLOCK
```

```

-----
ITEM  | TOTAL  NOT      OMIT  |      CATEGORIES
      |        PRESENT      |      1      2      3
-----
I001  |        |
  FREQ. |  500    0    0 |  152    13    335
  PERC. |        0.0  0.0 |  30.4    2.6    67.0
-----

```

2 SUBGROUP: FEMALE

BLOCK NO.: 1 NAME: BLOCK

ITEM	TOTAL	NOT PRESENT	OMIT	CATEGORIES		
				1	2	3
I001						
FREQ.	500	0	0	69	6	425
PERC.		0.0	0.0	13.8	1.2	85.0

TOTAL

BLOCK NO.: 1 NAME: BLOCK

ITEM	TOTAL	NOT PRESENT	OMIT	CATEGORIES		
				1	2	3
I001						
FREQ.	1000	0	0	221	19	760
PERC.		0.0	0.0	22.1	1.9	76.0

Item means, initial slope estimates, and Pearson and polyserial item-test correlations are given in the next table. For a detailed discussion of the measures shown here, refer to the discussion of the Phase 1 output of Section 5.1.

BLOCK ITEM	RESPONSE MEAN S.D.*	TOTAL SCORE MEAN S.D.*	PEARSON & POLYSERIAL CORRELATION	INITIAL SLOPE	INITIAL LOCATION
BLOCK 1 I001	2.539 0.831*	13.162 3.765*	0.714 0.976	1.000	0.000
CATEGORY	SCORING	MEAN	S.D.	PARAMETER	
1	1.000	8.190	2.235	0.000	
2	2.000	11.263	2.899	-0.155	
3	3.000	14.655	2.735	1.596	

## Phase 2 output

For the DIF model, a separate prior distribution is used for each group member, and the prior distribution is updated after each estimation cycle based on the posterior distribution from the previous cycle.

For the DIF model, it is assumed that different groups have different distributions with mean  $\mu_g$  and standard deviation  $\sigma_g$ . The distributions are not necessarily normal. These empirical posterior distributions are estimated simultaneously with the estimation of the item parameters. To obtain those parameters, the following constraint is imposed for the DIF model:

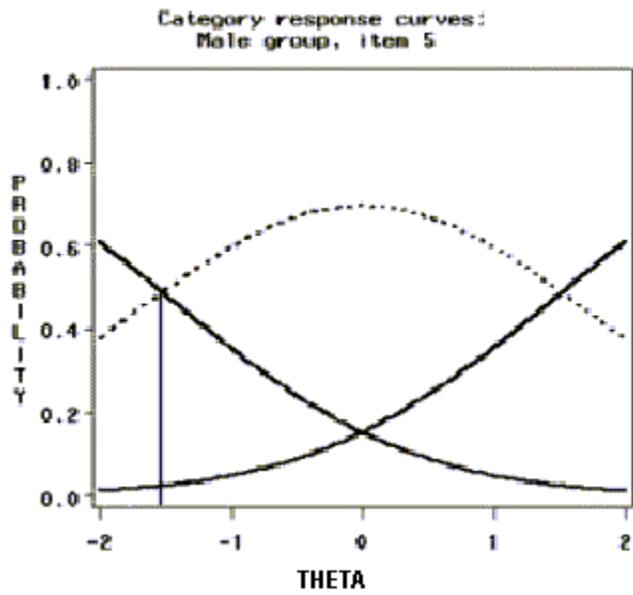
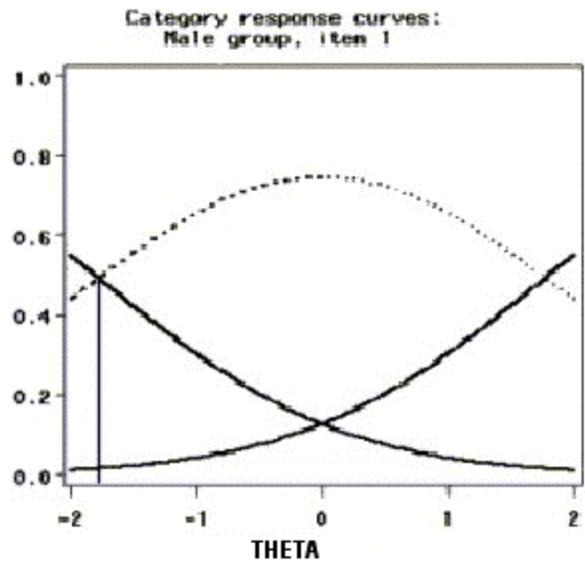
$$\sum_{j=1}^J d_{R_j} = \sum_{j=1}^J d_{F_j}.$$

This constraint implies the overall difficulty levels of a test or a set of common items given to both the reference group and focal group, indicated by subscripts  $R$  and  $F$ , respectively, are the same. Therefore, the item difficulty parameters for the focal groups are adjusted. Any overall difference in terms of test difficulty will be assumed to be the difference in ability level for subgroups. The ability level difference among groups can then be estimated by the posterior distributions.

The first difference between the output file discussed here and the Phase 2 output for Section 5.1 concerns the scoring function and step parameters for the multiple blocks. As no scoring function was specified on the CALIB command, the default scoring function 1, 2 will be used.

Under the partial credit model, the step parameters, also known as the item step difficulties or category intersections, correspond to the points on the ability scale where two successive item response category characteristic curves (IRCCC) intersect. The increasing difficulty of a step relative to other steps within an item is associated with higher values of the step parameters. In this example, where each item has 3 categories, 2 “steps” are needed to move from the first category to the third category: a respondent needs to move from category 1 to category 2, and a second step is needed to move from category 2 to category 3. From the second step parameters of items 1 and 2 (see below) moving from category 2 to category 3 is harder to do in the case of item 2 for the male respondents.

The IRCCC for items 1 and 5 are shown below. Vertical lines were added to indicate the trait level at which the curves for step 0 and step 1 intersect. The most likely response for a male with trait level of -2 would be to complete 0 steps in both cases. For a male with trait level of approximately 1.5, completing the step from category 2 to category 3 would be more likely in the case of item 5. Although there is little difference between the two graphs, it would appear that completing the first step is somewhat easier for item 1 than for item 5, while completing the second step is easier for item 5. This is in agreement with the second step parameters for these items: 1.769 for item 1 and 1.517 for item 5.



MULTIPLE GROUP MODEL [DIF (TREND) MODEL]

[GROUP: 1 MALE ]

ITEM BLOCK	1	BLOCK			
SCORING FUNCTION	:	1.000	2.000	3.000	
STEP PARAMTER	:	0.000	-1.743	1.743	
S.E.	:	0.000	0.168	0.163	

```

ITEM BLOCK 2 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.733 1.733
S.E. : 0.000 0.155 0.156

ITEM BLOCK 3 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.733 1.733
S.E. : 0.000 0.146 0.140

ITEM BLOCK 4 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.755 1.755
S.E. : 0.000 0.149 0.147

ITEM BLOCK 5 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.658 1.658
S.E. : 0.000 0.171 0.154

ITEM BLOCK 6 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.974 1.974
S.E. : 0.000 0.177 0.186

```

The step parameter information is followed by the item parameter estimates for the male group. Standard errors are computed from the empirical information matrix in the final Newton cycle.

```

+-----+-----+-----+-----+-----+-----+-----+-----+
| ITEM | BLOCK | SLOPE | S.E. | LOCATION | S.E. | GUESSING | S.E. |
+-----+-----+-----+-----+-----+-----+-----+-----+
| I001 | 1 | 0.846 | 0.054 | -0.590 | 0.070 | 0.000 | 0.000 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| I002 | 2 | 0.948 | 0.060 | 0.519 | 0.066 | 0.000 | 0.000 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| I003 | 3 | 0.628 | 0.034 | -0.542 | 0.076 | 0.000 | 0.000 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| I004 | 4 | 0.615 | 0.034 | 0.544 | 0.077 | 0.000 | 0.000 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| I005 | 5 | 0.414 | 0.025 | -0.666 | 0.098 | 0.000 | 0.000 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| I006 | 6 | 0.344 | 0.021 | 0.658 | 0.110 | 0.000 | 0.000 |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Similar information for the female group is given next. Note that the slope for each item is common across the two groups. This implies that the same item discrimination is assumed over the groups.

[GROUP: 2 FEMALE ]

```

ITEM BLOCK 1 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.743 1.743
S.E. : 0.000 0.168 0.163

```

```

ITEM BLOCK 2 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.733 1.733
S.E. : 0.000 0.155 0.156

ITEM BLOCK 3 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.773 1.773
S.E. : 0.000 0.146 0.140

ITEM BLOCK 4 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.755 1.755
S.E. : 0.000 0.149 0.147

ITEM BLOCK 5 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.658 1.658
S.E. : 0.000 0.171 0.154

ITEM BLOCK 6 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
STEP PARAMTER : 0.000 -1.974 1.974
S.E. : 0.000 0.177 0.186

```

ITEM	BLOCK	SLOPE	S.E.	LOCATION	S.E.	GUESSING	S.E.
I001	1	0.846	0.054	-0.615	0.085	0.000	0.000
I002	2	0.948	0.060	0.644	0.057	0.000	0.000
I003	3	0.628	0.034	0.010	0.075	0.000	0.000
I004	4	0.615	0.034	-0.348	0.084	0.000	0.000
I005	5	0.414	0.025	-0.645	0.118	0.000	0.000
I006	6	0.344	0.021	0.877	0.098	0.000	0.000

DIF contrasts are given next. In the table below, the CONTRAST column gives the differences between item locations between the groups and the associated standard error. The STD column contains standardized contrasts, obtained by dividing each contrast by its standard error. The probability that a normal variate exceeds the absolute value of the standardized difference is also given. This is a one-sided test.

```

CONTRAST OF ITEM LOCATIONS:
GROUP 2: FEMALE MINUS REFERENCE GROUP 1: MALE
+-----+-----+-----+-----+
|ITEM|BLOCK|CONTRAST|STD|
| | | (S.E.) | (PROB. ) |
+=====+=====+=====+=====+
|I001| 1 | -0.025 | -0.230 |
| | | ( 0.110) | ( 0.409) |
| | | | |
|I002| 2 | 0.125 | 1.433 |
| | | ( 0.087) | ( 0.076) |
| | | | |

```

I003	3	0.552	5.176
		( 0.107)	( 0.000)
I004	4	-0.892	-7.798
		( 0.114)	( 0.000)
I005	5	0.021	0.138
		( 0.153)	( 0.445)
I006	6	0.219	1.494
		( 0.147)	( 0.068)

$\chi^2$ -test statistics for the item location contrasts are given in the next section of the output file. In this case, with only one degree of freedom,  $\chi^2 = (std. difference)^2$ . This is a two-sided test. In this table, these  $\chi^2$ -test statistics and exceedance probabilities are summarized.

CHI-SQUARES OF ITEM LOCATION CONTRASTS:

ITEM	BLOCK	CHI-SQRS	D.F. PROB.
I001	1	0.053	1. 0.803
I002	2	2.052	1. 0.148
I003	3	26.789	1. 0.000
I004	4	60.814	1. 0.000
I005	5	0.019	1. 0.861
I006	6	2.231	1. 0.131
TOTAL		91.958	6. 0.000

When the summary statistics for the 2 groups are compared, we see that only the standard deviation of the threshold differs. Recall that for this example, the DIF keyword on the MGROUP command was used to allow only threshold parameters to differ between the groups. Overall, no large difference between groups over all items is observed.

SUMMARY STATISTICS OF PARAMETER ESTIMATES

1 GROUP NAME: MALE

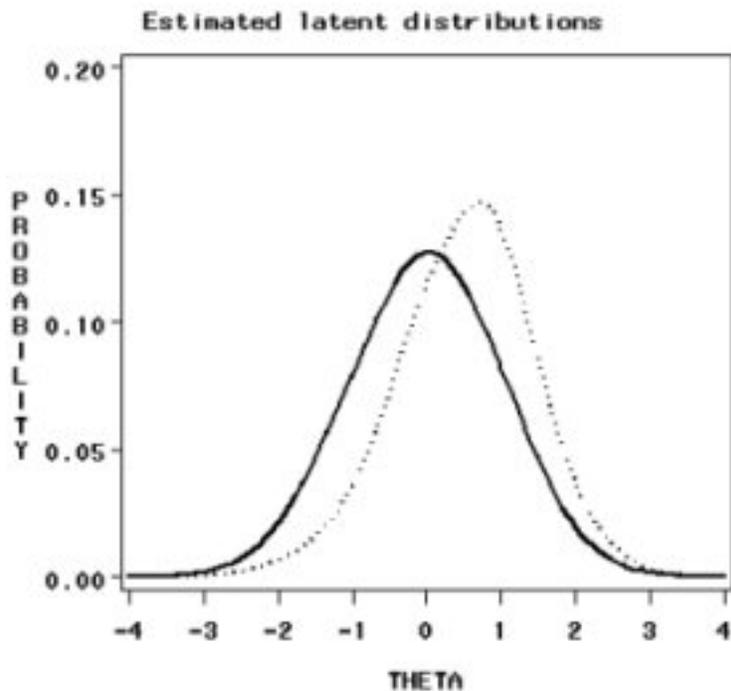
PARAMETER	MEAN	STN DEV	N
SLOPE	0.633	0.235	6
LOG(SLOPE)	-0.520	0.394	6
THRESHOLD	-0.013	0.645	6
GUESSING	0.000	0.000	0

2 GROUP NAME: FEMALE

PARAMETER	MEAN	STN DEV	N
SLOPE	0.633	0.235	6
LOG(SLOPE)	-0.520	0.394	6
THRESHOLD	-0.013	0.648	6
GUESSING	0.000	0.000	0

The final output is the estimated latent distributions by group. The origin and unit of the scale are set so that the mean and standard deviation of the reference group are 0 and 1 respectively.

A plot of the estimated latent distributions is given below. The solid line represents the distribution for the male group. If there is appreciable DIF, the latent distributions do not represent the same latent variable and no meaningful comparison of the two distributions is possible. If there is no DIF, significant differences between the latent distributions represent real differences between the populations sampled.



QUADRATURE POINTS AND POSTERIOR WEIGHTS:

GROUP 1 GROUP NAME: MALE

	1	2	3	4	5
POINT	-0.4000E+01	-0.3667E+01	-0.3333E+01	-0.3000E+01	-0.2667E+01
WEIGHT	0.5988E-04	0.2137E-03	0.6808E-03	0.1934E-02	0.4887E-02
	6	7	8	9	10
POINT	-0.2333E+01	-0.2000E+01	-0.1667E+01	-0.1333E+01	-0.1000E+01
WEIGHT	0.1096E-01	0.2172E-01	0.3790E-01	0.5826E-01	0.8000E-01
	11	12	13	14	15
POINT	-0.6667E+00	-0.3333E+00	0.3331E-15	0.3333E+00	0.6667E+00
WEIGHT	0.1009E+00	0.1178E+00	0.1249E+00	0.1190E+00	0.1034E+00
	16	17	18	19	20
POINT	0.1000E+01	0.1333E+01	0.1667E+01	0.2000E+01	0.2333E+01
WEIGHT	0.8257E-01	0.5917E-01	0.3742E-01	0.2086E-01	0.1029E-02
	21	22	23	24	25
POINT	0.2667E+01	0.3000E+01	0.3333E+01	0.3667E+01	0.4000E+01
WEIGHT	0.4516E-02	0.1766E-02	0.6167E-03	0.1924E-03	0.5368E-04

TOTAL WEIGHT: 1.00000  
 MEAN : 0.00000  
 S.D. : 0.99974

QUADRATURE POINTS AND POSTERIOR WEIGHTS:

GROUP 2 GROUP NAME: FEMALE

	1	2	3	4	5
POINT	-0.4000E+01	-0.3667E+01	-0.3333E+01	-0.3000E+01	-0.2667E+01
WEIGHT	0.1485E-04	0.5381E-04	0.1748E-03	0.5093E-03	0.1331E-02
	6	7	8	9	10
POINT	-0.2333E+01	-0.2000E+01	-0.1667E+01	-0.1333E+01	-0.1000E+01
WEIGHT	0.3120E-02	0.6569E-02	0.1248E-01	0.2175E-01	0.3608E-01
	11	12	13	14	15
POINT	-0.6667E+00	-0.3333E+00	0.3331E-15	0.3333E+00	0.6667E+00
WEIGHT	0.5834E-01	0.8712E-01	0.1130E+00	0.1320E+00	0.1437E+00
	16	17	18	19	20
POINT	0.1000E+01	0.1333E+01	0.1667E+01	0.2000E+01	0.2333E+01
WEIGHT	0.1360E+00	0.1059E+00	0.6927E-01	0.3922E-01	0.1955E-01
	21	22	23	24	25
POINT	0.2667E+01	0.3000E+01	0.3333E+01	0.3667E+01	0.4000E+01
WEIGHT	0.8653E-02	0.3410E-02	0.1199E-02	0.3764E-03	0.1056E-04

TOTAL WEIGHT: 1.00000  
 MEAN : 0.00000  
 S.D. : 0.99974