

Paper PO-090

Two SAS® Macros for DIF Analysis

Shiqi Hao

Abstract

Detection of differential item functioning (DIF) is a common procedure used in testing and measurement. Two SAS macros are introduced that can produce the DIF severity classification indexes for dichotomously scored items and for polytomously scored items based on the National Assessment of Educational Progress (NAEP) and Educational Testing Service (ETS) guidelines. Detailed steps and procedures from reading raw data to output the result in Excel are discussed.

Key words: DIF, multiple-choice, constructed-response, Mantel-Haenszel statistics, standardized mean difference

Introduction

Detection of Differential item functioning (DIF) is a frequently used procedure in large-scale testing and measurement as it is related to fairness and validity of a test or measure. DIF analyses are primarily designed to detect differential item performance across subgroups of populations while controlling for ability. For example, a group of high-ability students are expected to have better performance on an item than a group of low-ability students. All things considered, this performance is the desired result in a “fair” testing environment. However, DIF procedures seeks to identify those items where subgroups of students (e.g., male or female, racial/ethnic groups) with similar ability perform at measurably different levels. For this analysis, subgroups were assigned to either the reference or focal groups. It is a common practice to assign males and whites to the reference group and females and African-Americans to the focal group. In addition, all students were placed in one of a fixed number (e.g., eight or ten) ability groups according to their total test scale score, the result being samples from these ability groups of approximately the same size.

For the multiple-choice items, DIF is assessed most often by utilizing the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959). The MH procedure computes an odds-ratio statistic by comparing the observed and expected totals of an $8 \times 2 \times 2$ contingency table. For example, each of the eight ability groups has its own 2×2 contingency table, formed by a cell for each type of response, correct and incorrect, within each focal group.

For the polytomous items, DIF is assessed by examining the standardized mean difference (SMD), utilizing a procedure put forth by Dorans, Schmitt, and Bleistein (1992). The SMD is the difference between the unweighted item mean of the focal group and the weighted item mean of the reference group. The weights applied to the reference group are applied so that the

weighted number of reference group students is the same as in the focal group (within the same ability group). The SMD is then divided by the total group item standard deviation, resulting in a measure of the effect size (ES) for the SMD.

Both the dichotomous and polytomous item DIF statistics were placed in severity classifications based on National Assessment of Educational Progress (NAEP) and Educational Testing Service (ETS) guidelines for the dichotomous items and on NAEP guidelines for the polytomous items (Allen, Carlson, & Zalanak, 1999). This paper intends to discuss two SAS macros that are used to detect the DIF for dichotomously scored items and for polytomously scored items.

We first discuss the macro for multiple-choice or dichotomously scored items and then the macro for polytomously scored items.

Part I Macro 1 for Dichotomous Item DIF

Based on NEAP and ETS criteria, the dichotomous DIF classification is defined as follows:

Rule 1: MH delta (MHD) not significantly different from 0 (based on $\alpha=.05$) **or**

$|MHD| < 1.0$ are classified as A. [Note: The MHD is the ETS delta scale for item difficulty where the natural logarithm of the common odds ratio is multiplied by $-(4/1.7)$.]

Rule 2: MHD significantly different from 0 and $\{|MHD| \geq 1.0 \text{ and } < 1.5\}$ **or**

MHD not significantly different from 0 and $|MHD| \geq 1.0$ are classified as B.

Rule 3: $|MHD| \geq 1.5$ and significantly different from 0 are classified as C.

SAS codes:

Step 1: Reading the raw data and scoring the responses into dichotomous scores (e.g., either 1 for right or 0 for wrong). This step can be skipped if dichotomous scores are already obtained.

Step 2: Computing MH odds ratio and 95% confidence intervals of the DIF statistics

The following macro uses the gender groups as an example for a 50 dichotomously scored item test.

```
-----
options mprint mlogic spool; /*****computing MH odds ratio and 95% confidence limits and
outputting them into separate data files***/
%macro gdif(num); **num=number of items;
%do i=1 %to &num;
title " Gender DIF variable is S&i";
proc rank data=DIF2 out=Cdata groups=10;
var rawscore; /***using the total score on the test as control variable***/
ranks stratum;
proc freq data=Cdata noprint; /**noprint option prevents printing out all the MH statistics**/
tables stratum*Gender*S&i/CMH norow nocol nopercnt ;
```

```

output out=out&i lgor mhor;
run;
%end;
%put &i;
%mend;
%gdiff(50); /**make necessary change for item numbers**/
    data all (RENAME=(_MHOR_=md L_MHOR=lower U_MHOR=upper));
    set
    %do i = 1 %to &num;
        out&i /*(drop =_LGOR_ L_LGOR U_LGOR); */
    %end;
run;

/*Pulling out MH table values and computing the ETS Delta classes and exporting to EXCEL*/
    data final; set all;
    item=_N_;
    const = -2.3529;
    mhd=const*(log(md));
    lowerd=const*(log(lower));
    upperd=const*(log(upper));
    /*******
    IF (ABS(mhd) < 1.0) or (lowerd< 0 <upperd) or (upperd<0<lowerd) then
    class= "A" ;
    IF (1.5 >ABS(mhd)>= 1.0 and lowerd< 0>upperd) or (1.5 >ABS(mhd)>=
    1.0 and (lowerd> 0 <upperd)) or ((upperd<0<lowerd)
    and (ABS( mhd) >= 1.0)) or ((upperd>0>lowerd) and ABS( mhd) >= 1.0)
    then class = "B";
    IF (ABS(mhd) >= 1.5 and lowerd< 0>upperd) or (ABS(mhd) >= 1.5
    and (upperd>0<lowerd)) then class = "C";
    *****/
    ***the following adds the negative or positive sign for the DIF
    classification;
    IF (ABS(mhd) < 1.0) and mhd <0 then class ="A-";
    ELSE IF (ABS(mhd) < 1.0) and mhd >0 then class ="A+";
    If lowerd< 0 <upperd and mhd <0 then class ="A-";
    ELSE If lowerd< 0 <upperd and mhd >0 then class ="A+";
    If upperd<0<lowerd and mhd <0 then class ="A-";
    ELSE If upperd<0<lowerd and mhd >0 then class ="A+";

```

```

IF 1.5 >ABS(mhd)>= 1.0 and lowerd< 0>upperd and mhd > 0 then
class="B+";
ELSE IF 1.5 >ABS(mhd)>= 1.0 and lowerd< 0>upperd and mhd < 0 then
class="B-";
If 1.5 >ABS(mhd)>= 1.0 and lowerd> 0 <upperd and mhd >0 then class="B+";
ELSE If 1.5 >ABS(mhd)>= 1.0 and lowerd> 0 <upperd and mhd <0 then
class="B-";
If upperd<0<lowerd and ABS( mhd) >= 1.0 and mhd >0 then class="B+";
ELSE If upperd<0<lowerd and ABS( mhd) >= 1.0 and mhd <0 then
class="B-";
If upperd>0>lowerd and ABS( mhd) >= 1.0 and mhd >0 then class="B+";
ELSE If upperd>0>lowerd and ABS( mhd) >= 1.0 and mhd <0 then
class="B-";
IF ABS(mhd) >= 1.5 and lowerd< 0>upperd and mhd >0 then class="C+";
ELSE IF ABS(mhd) >= 1.5 and lowerd< 0>upperd and mhd <0 then
class="C-";
If ABS(mhd) >= 1.5 and upperd>0<lowerd and mhd >0 then class="C+";
ELSE If ABS(mhd) >= 1.5 and upperd>0<lowerd and mhd <0 then
class="C-";
keep item md lower upper const mhd lowerd upperd class;
run;
proc print data=final; run;
%mend;
%gdif(50);

```

Step 3 Exporting the DIF classification to Excel

There are several ways of exporting SAS outputs into Excel external file. The following are some examples.

Example 1:

```

ods listing close; /****output option1--outputting the file into an Excel file via xml*****/
ods tagsets.msoffice2k file='C:\...a72dif.xls';
proc print data=final;
var item md lower upper const mhd lowerd upperd class; run;
ods tagsets.msoffice2k close;
ods listing;

```

Example 2:

```
/*****output option2--when direct Xls output doesn't work*****/  
proc export data=final outfile='C:\...\a72dif.txt' DBMS=DLM REPLACE; DELIMITER=' ';  
run;
```

Example 3:

```
/*****option 3 using "file" and "put" to produce a new external data set**/  
data final; set final;  
      /*****you can rearrange the order of variables****/  
file 'C:\...\DIF output\A70dif.txt' dlm= ',' lrecl=2500 ;  
put  
  item $ 3-10  
    @13 md 7.4  
    @35 lower 7.4  
    @55 upper 7.4  
    @65 const 7.4  
    @85 mhd 7.4  
    @95 lowerd 7.4  
    @115 upperd 7.4  
    @125 class $ 2.0  
  ;  
run;
```

Part II Macro 2 for Polytomous Item DIF

Based on ETS and NEAP guidelines, the polytomous DIF classification is defined as follows:

Rule 1: If the p-prob is $>.05$, classify the item as AA.

Otherwise: Rule 2: If $|ES| \leq .17$, classify as AA.

Rule 3: If $|ES|$ is $>.17$ but $\leq .25$, classify as BB.

Rule 4: If $|ES|$ is $>.25$, classify as CC.

SAS Codes:

Step 1; Reading from raw data and producing the ratings or ordinal scores. (Skipped)

Step 2: computing MH odds ratio and 95% confidence limits and then outputting them into separate data files

```

options mprint mlogic spool; /****computing MH odds ratio and 95% confidence limits and
outputting them into separate data files***/
%macro gdif(num);
%do i=1 %to &num;
title " Gender DIF variable is elaew&i";
proc rank data=DIF2 out=Cdata groups=3;
var raw; /***using the total score on the test as control variable***/
ranks stratum;
data Cdata; set Cdata; if stratum eq . then delete;
proc freq data=Cdata noprint; /**noprint option prevents printing out all the MH statistics**/
tables stratum*sexe*elaew&i/CMH nocol norow nopercnt;
title 'MH for Combined Group';
output out=out&i cmhga;
proc means data=DIF2 mean std sumwgt maxdec=3; var elaew&i;
output out=comb&i mean=cbmean n=cbnumwgt std=cbstd;
title 'Mean for combined group';
run;

```

```

data focal; set Cdata; if sexe='F'; *make necessary change for ethnicity focal group;
proc means data=focal maxdec=3 mean sumwgt; var elaew&i; output out=focalm&i
mean=focmean std=fostd n=focnumwgt; run;
title 'Mean for focal group';
run;

```

```

ods trace on/listing;
proc freq data=focal;
tables stratum/nocum out=focfreq&i;
title 'Freq for focal group';
run;
ods trace off;

```

```

Data ref; set Cdata; if sexe='M'; **for reference group;
proc sort data=ref; by stratum;
proc means data=ref maxdec=3 mean sumwgt; var elaew&i; by stratum; output out=refm&i
mean=refmean std=refstd;
title 'Mean for Reference Group';
run;

```

```

%end;
%put &i;
%mend;
%gdif(4); /**make necessary change for item numbers**/

/**Step 3: Calculate the standardized mean difference**/
data twoa; set out1 out2 out3 out4; item=_n_; proc print; run;
data twob; set comb1 comb2 comb3 comb4; item=_n_; proc print; run;
data twoc; set focalm1 focalm2 focalm3 focalm4; item=_n_; proc print; run;
data twod; set refm1 refm2 refm3 refm4; proc print; run;
data twoe; set focfreq1 focfreq2 focfreq3 focfreq4 ;proc print; run;
data refsmean;
if stratum eq . then delete;
set twod; set twoe ;
refsmean=(PERCENT*refmean)/100;
proc print data=refsmean; sum refsmean;
run;
proc sort data=refsmean; by stratum;
proc transpose data=refsmean out=transposed;
by stratum; var refsmean;
run;

proc transpose data=transposed out=stmeanr; by _name_;
var col1-col4; /**col1-col4 stands for number of items**/
run;
proc print data=stmeanr;
run;
data refmeanfinal ; set stmeanr;
item=_N_;
standrefmean=sum (of col1-col4); /**col1--col4 now stands for ability interval or stratum*/
run;

/**Step 4: Classifying the items into DIF classes**/
%let pvalue=p_cmhga; /**use the general association statistics**/
%let df=df_cmhga;
data final; set twoa;
      set twob;

```

```

        set twoc;
        set twod;
        set twoe;
        set refmeanfinal; by item;
        effectsize=(focmean-standrefmean)/cbstd;
        if &pvalue >.05 then class="AA";
        else if abs(effectsize)>. or abs(effectsize) le .17 then class="AA";
        else if abs(effectsize)>.17 and abs(effectsize) le .25 then
        class="BB"; else if abs(effectsize)>.25 then class="CC";
        proc print data=final;
            run;
%macro classdif(num, col);
%let pvalue=p_cmhga; /**use the general association statistics**/
%let df=df_cmhga;
%do i=1 %to &num;
%do i=1 %to &col;
data refsdmean;
*if stratum eq . then delete;
set twod; set twoe ;
refsdmean=(PERCENT*refmean)/100;
proc print data=refsdmean; sum refsdmean;
run;
proc sort data=refsdmean; by stratum;
proc transpose data=refsdmean out=transposed;
by stratum; var refsdmean;
run;
proc print data=transposed; sum col&i;
run;

proc transpose data=transposed out=stmeanr; by _name_;
var col&i; /**col1-col1 stands for number of items**/
run;
proc print data=stmeanr;
run;
data refmeanfinal ; set stmeanr;
item=_N_;
standrefmean=sum (of col&i); /**col1--coli now stands for ability interval or stratum*/

```

```
keep standrefmean item;  
run;
```

```
data final; set twoa; set twob; set twoc; set refmeanfinal;
```

```
item=_N_;
```

```
smd=focmean-standrefmean;
```

```
es=smd/cbstd;
```

```
run;
```

```
proc print data= final;
```

```
run;
```

```
proc print data=out&i; run;
```

```
%end;
```

```
%end;
```

```
%put &i;
```

```
%mend classdif;
```

```
%classdif (num=4, col=4);
```

```
run;
```

```
/**Step 5: Merging the separate data files and exporting into Excel**/
```

```
Example 1:
```

```
data etseffect;
```

```
length item $3.
```

```
merge out&i focalm&i;
```

```
proc export data=final outfile='C: \DIF output\dif.txt' DBMS=DLM REPLACE; DELIMITER=' ';
```

```
run;
```

```
Example 2:
```

```
data final; set final; /******option 3 using "file" and "put" to produce a new external data set**/
```

```
/******you can rearrange the order of variables****/
```

```
file 'C: \DIF output\dif.txt' dlm= ' ' lrecl=2500 ;
```

```
put
```

```
item $ 3-10
```

```
@13 md 7.4
```

```
@35 lower 7.4
```

```
@55 upper 7.4
```

```
@65 const 7.4
```

```
@85 mhd 7.4
```

```
@95 lowerd 7.4
```

```
@115 upperd 7.4  
@125 class $ 2.0;  
run;
```

Conclusion Remarks

As DIF is a common procedure in testing and measurement, particularly in large-scale testing, it has become one aspect of the necessary analyses in testing to ensure test fairness and validity. The two SAS macros discussed above are efforts to facilitate this kind of data analysis via SAS and output the classification results into Excel or other reporting platforms. The author welcomes any suggestions and comments on further improvement.

References

- Allen, N. L., Carlson, J. E., & Zalanak, C. A. 1999. *The NAEP 1996 Technical Report*. Washington, DC: National Center for Education Statistics.
- Clauser, Brian E., & Mazor, K. M. (1998). An NCME Instructional Module on Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice* 17, no. 1 (spring): 31–44.
- Dorans, N. J., Schmitt, A. P., & Bleistein, C.A. (1992). The Standardization Approach to Assessing Comprehensive Differential Item Functioning. *Journal of Educational Measurement* 29:309–19.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Notes: I thank Dr. Christina Schneider and Melinda Taylor (my former colleagues at SC Department of Education), and Prof. Keh-Shin Lii at the University of California, Riverside, Dr. Shuqin Guo at California Department of Education and an anonymous reviewer from ETS for their generous support and/or valuable comments in the development of the two SAS programs.

CONTACTS

Shiqi Hao, University of Massachusetts Boston, Center of Science and Math in Context (COSMIC), 100 Morrissey Blvd, Boston, MA02125. Email: Shiqi.Hao@umb.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.