Using the **difR** Package with Lertap 5

Larry R Nelson[1]
Curtin University, Western Australia
www.lertap.com

Document date: 24 January 2019

difR is the name of an R package designed to support "DIF" analyses, where DIF=differential item functioning. It is introduced at this webpage. When I undertook the work described below, difR was into Version 5, published 2018-05-14.

The use of R packages with Lertap 5 is discussed in this document. I will use install.packages("difR") for my work in this paper.

Lertap 5 has its own DIF analysis capability, based on the Mantel-Haenszel (M-H) method. The "Ibreaks" routine in Lertap is the gateway to M-H with Lertap. A technical paper discussing and exemplifying the use of Lertap M-H is here.

The **difR** package has more extensive support for DIF researchers, featuring the M-H method and several others, including logistic regression and SIBTEST (to name just two of the others – there are more).

This paper is for those who have item response data in a Lertap 5 workbook and would like to use the data with the **difR** package.

When test results are input to difR, it is expected that the input will consist of item scores and a group variable. Item scores will be zero and one (0,1) – these are often referred to as "binary" scores. The group variable is expected to have two categories[2]. It might, for example, be Gender with codes of F and M, or, as in the example below, Race with codes of B and W. The version of difR used for this paper would allow for missing data among the item scores, but not in the group variable. The codes used for the group variable could be letters or digits.

To prepare their data for difR, Lertap users will have to get Lertap to make item scores, make sure that their group variable has just two codes, make sure that all data records have a group code (no missing data), and then prepare a "csv" file.

My suggestion is that the csv file start with columns with the items scores, one column per test item, and end with the group code in the last column. The first row in the csv file will be for variable "headers". Item headers will typically be, for example, Item1 or Q1.

Creating a csv file from a Lertap 5 workbook

I have selected the "exam1" test results from this example. Exam1 is used as a major exhibit in Patrick Meyer's 2014 text "Applied Measurement with jMetrik" (see the Lertap references).   I have used these results in my "GimmeABreak" paper, where they featured in Example 2 (pp. 15-19).

---

[1] Comments / questions may be sent to l.nelson@curtin.edu.au
[2] difR supports more than two groups.

The screen snapshot above displays some of the rows and columns in the Lertap 5 Data worksheet corresponding to Meyer's "exam1" test results. In this case, the three initial columns were used to house descriptive information; these columns were followed by item responses. There were a total of 56 multiple-choice items; their responses were found starting in column 4, ending in column 59.

I will need to have item scores before I'm able to use the difR routines. The screen shot above shows item *responses*, not item *scores*. To get item scores, I'll use the "Item scores and correlations" option in Lertap.

The "Item scores and correlations" option adds the "IStats" worksheet to the workbook. It has two sections: item scores, followed by item statistics and correlations.

Next, I take the "BinaryItems" option in Lertap. It creates a new workbook with item scores – here's how part of it looks at this stage[3]:

---

[3] Excel has named the new workbook "Book8" in this case – the book number will vary.

Notice that the BinaryItems option has not copied the Gender and Race infor-
mation from the original workbook. As I'd like to look at results based on race, I
now need to copy the race column from the original workbook and paste it into
the new workbook's Data sheet. When I do so, I paste it on the right-hand side
of the new workbook:

Using the difR package, page 3.

Next I need to delete the first row so that what shows as row 2 below will become the first row. I will not need the id column, and will also delete it.

	1	2	3	4	5	6	7	8	9
1	Lertap5 IScores matrix, created: 23/1/19								h
2	id	item1	item2	item3	item4	item5	item6	item7	item8
3	1		1.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00

I now tell Excel to save this new workbook, giving it an appropriate name: "Exam1RaceItemScores-BW-only.xlsx".

Exam1RaceItemScores-BW-only.xlsx - Excel   Larry Nelson

File   Hom Inser Draw Page Form Data Revi View Help   Tell me

AutoSave Off   Gridlines   Formula Bar

	1	2	3	4	53	54	55	56	57
1	item1	item2	item3	item4	item53	item54	item55	item56	race
2	1.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	W
3	1.00	0.00	1.00	0.00	1.00	1.00	1.00	0.00	W
4	1.00	0.00	0.00	1.00	1.00	0.00	1.00	1.00	W
5	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	W
6	1.00	1.00	1.00	1.00	0.00	1.00	0.00	1.00	W
7	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	W
8	1.00	0.00	1.00	1.00	1.00	0.00	0.00	1.00	W
9	1.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	W
10	1.00	1.00	1.00	0.00	0.00	0.00	1.00	1.00	W
11	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	W
12	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	W
13	1.00	1.00	1.00	0.00	1.00	0.00	0.00	1.00	W

Data   CCs   +

The snapshot above shows that the new workbook has **item1** scores in the first column, then more item scores, and, at the end, **race** in column 57. This exemplifies the format for a workbook that will best serve my needs when using **difR**: columns with item scores followed by the "group" variable for DIF (which is race in the case).

Now, with the focus on the Data sheet (as in the snapshot above), I ask Excel to save as a "CSV (MS-DOS)" file, highlighted in blue below:
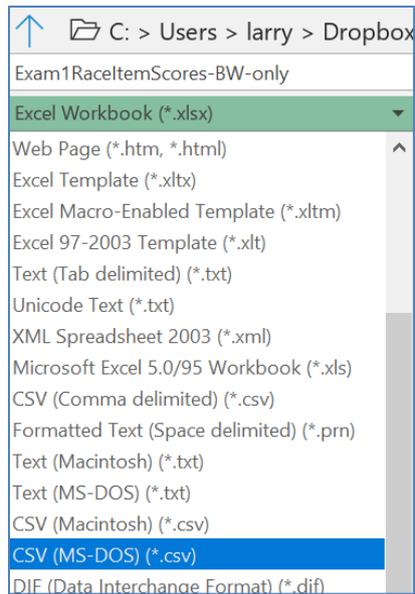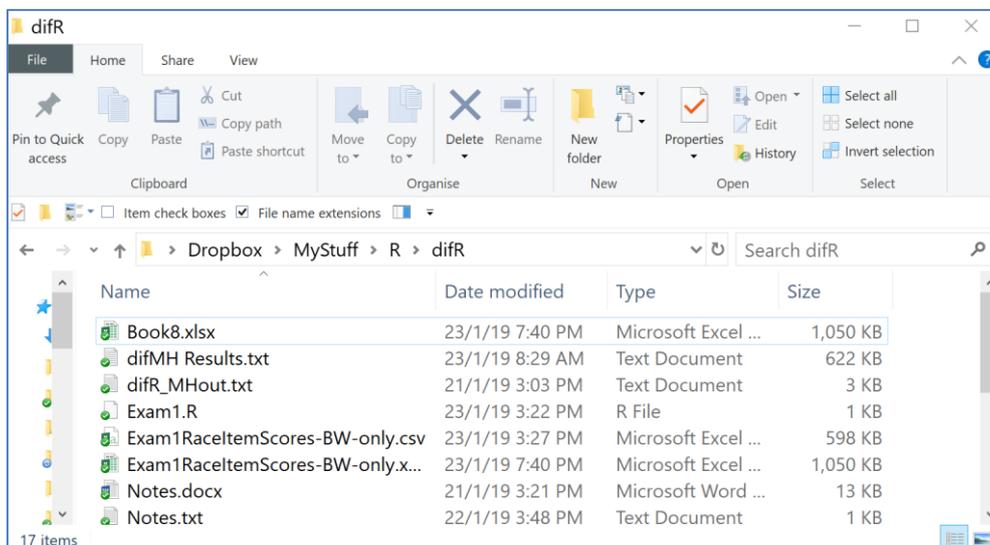
Excel sounds a warning:
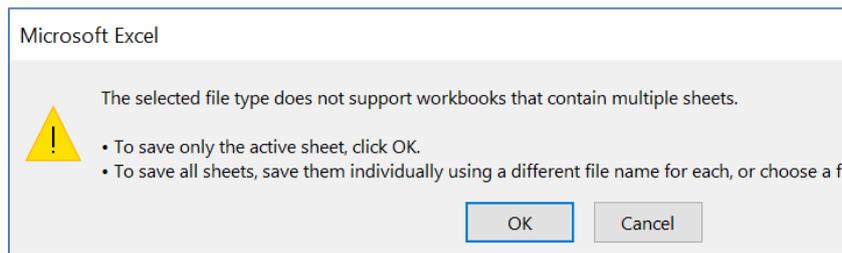
I click OK and close the workbook.

By using the Windows file manager I can confirm that it has indeed been saved as a file named "Exam1RaceItemScores-BW-only.csv".

This is how things should be at this stage. I now have a csv file that can readily be used with the **difR** package.

M-H results with difR

I have both R and RStudio installed on my laptop, as mentioned in this paper. Were I to continue using **difR**, and well I might, I'd probably develop a script for use with RStudio. For the moment, however, I'll just run with R.

The "Exam1.R" script file shows in the file list above. Here's the coding:

```
# Made to work with the Exam1 dataset, 24 January 2019
# By Larry Nelson, Curtin University, Western Australia
# l.nelson@curtin.edu.au
setwd("C:/Users/larry/Dropbox/MyStuff/R/difR/Exam1")
library(difR)
dir()
AAitems <-read.csv(file="Exam1RaceItemScores-BW-only.csv")
difItems<-AAitems[colnames(AAitems)]
head(difItems)
difMH(difItems, group="race", focal.name="W", save.output=TRUE)
```

When R finishes running the code lines, it provides this message:

```
Output was captured and saved into file
 'C:/Users/larry/Dropbox/MyStuff/R/difR/Exam1/out.txt'
```

Some of the output is shown below:

```
Mantel-Haenszel Chi-square statistic:

        Stat.    P-value
item1   19.3862  0.0000 ***
item2   14.5016  0.0001 ***
item3    7.1053  0.0077 **
item4   21.0385  0.0000 ***
item5    1.4150  0.2342
item6   17.1696  0.0000 ***
item7    0.1694  0.6806
item8    0.9321  0.3343
item9    6.5306  0.0106 *
item10   0.0695  0.7920


Items detected as DIF items:

 item1
 item2
 item3
 item4
 item6
 item9
 item16
```
 (there were 25 items in this list, all with P-values less than .05)

```
Effect size (ETS Delta scale):

Effect size code:
 'A': negligible effect
 'B': moderate effect
 'C': large effect
```

```
        alphaMH deltaMH
item1   0.7198  0.7726 A
item2   0.7791  0.5865 A
item3   0.7722  0.6076 A
item4   0.7396  0.7090 A
item5   0.9179  0.2014 A
item6   1.4001 -0.7909 A
item7   1.0295 -0.0683 A
item8   0.8904  0.2728 A
item9   0.8404  0.4085 A
item10  1.0327 -0.0756 A

item37  2.0179 -1.6499 C

item45  1.8638 -1.4632 B
```

Only two items, item37 and item45, evidenced practical significance as indexed by the deltaMH statistic.

As I mentioned earlier, difR can look for possible DIF using numerous methods – here I've used only one, Mantel-Haenszel, the method used in Lertap 5, and a method with good support in the literature.

difR is able to make graphical summaries of DIF.

Have a look at these URLs:

https://cran.r-project.org/web/packages/difR/difR.pdf

https://eeecon.uibk.ac.at/psychoco/2011/slides/Magis_hdt.pdf