
Code ▾

Item-Focussed Trees for Detecting Differential Item Functioning (DIF) using Logistic Regression

2022-10-19

Data Source

Real-world data to demonstrate the implementation of the methods discussed in this workshop were from the population-based Winnipeg Regional Health Authority (WRHA) Joint Replacement Registry (JRR), for patients having a total or partial hip or knee replacement.

We selected 1391 individuals who had a total hip replacement and provided complete responses to the SF-12 (version 2) physical and mental components items prior to undergoing surgery (i.e., at the baseline clinic visit).

Objective

To test for uniform and non-uniform DIF by age and sex on the physical and mental component items of the SF-12

Packages and Functions to Test for Tree-Based DIF in R

We would explore the **DIFtree** package.

Install the following packages, if not previously installed.

Hide

```
# install.packages(c("tidyverse", "DIFtree"))
```

Load the Libraries

Hide

```
library(tidyverse) # For data manipulation
library(DIFtree) # For tree-based DIF Test
library(DT) # Create tables
```

Load the dataset and carry out the necessary data manipulation

Hide

```

tha_data <- read_csv("SF12health_Complete_Hip.csv")

# glimpse(tha_data)

allitems_sex_age <- c("SF1_PH", "SF2A_PH", "SF2B_PH", "SF3A_PH",
                    "SF3B_PH", "SF4A_MH", "SF4B_MH", "SF5_PH",
                    "SF6A_MH", "SF6B_MH", "SF6C_MH", "SF7_MH", "Sex", "age")

allitems <- c("SF1_PH", "SF2A_PH", "SF2B_PH", "SF3A_PH",
             "SF3B_PH", "SF4A_MH", "SF4B_MH", "SF5_PH",
             "SF6A_MH", "SF6B_MH", "SF6C_MH", "SF7_MH")

tha_data_sx_ag <- tha_data[,allitems_sex_age]

```

Assumption: We assume each dimension (physical health & mental health) are unidimensional. Hence, we skip test(s) to assess dimensionality.

First, we need to create a separate data matrix including the item response columns. We are using logistic regression (LR). Therefore, we need to change the five and three category responses to two-category responses.

Hide

```
sf_items <- tha_data_sx_ag[,c(1:12)]
```

Select items associated with physical health and mental health

Hide

```

PH_sf_items <- sf_items[,c(1,2,3,4,5,8)]
#5-response category items: Re-coding item responses 1 and 2 as 0, and the remaining responses (3, 4, 5) as 1
#3-response category items: Re-coding item responses 1 as 0, and the remaining responses (2 and 3) as 1
PH_sf_items$SF1_PH <- ifelse(PH_sf_items$SF1_PH==1|PH_sf_items$SF1_PH==2, 0, 1)
PH_sf_items$SF2A_PH <- ifelse(PH_sf_items$SF2A_PH==2|PH_sf_items$SF2A_PH==3, 1, 0)
PH_sf_items$SF2B_PH <- ifelse(PH_sf_items$SF2B_PH==2|PH_sf_items$SF2B_PH==3, 1, 0)
PH_sf_items$SF3A_PH <- ifelse(PH_sf_items$SF3A_PH==1|PH_sf_items$SF3A_PH==2, 0, 1)
PH_sf_items$SF3B_PH <- ifelse(PH_sf_items$SF3B_PH==1|PH_sf_items$SF3B_PH==2, 0, 1)
PH_sf_items$SF5_PH <- ifelse(PH_sf_items$SF5_PH==1|PH_sf_items$SF5_PH==2, 0, 1)
PH_sf_items <- as.matrix(PH_sf_items)

MH_sf_items <- sf_items[,c(6,7,9,10,11,12)]
#Re-coding item responses 1 and 2 as 0 and the remaining responses (3, 4, 5) as 1
MH_sf_items <- ifelse(MH_sf_items==1|MH_sf_items==2, 0, 1)
MH_sf_items <- as.matrix(MH_sf_items)

```

Obtain the data frame including covariates (i.e., sex, age)

Hide

```
covariate<-tha_data_sx_ag[,c(13:14)]
covariate$Sex<-as.numeric(covariate$Sex)
covariate = data.frame(covariate)
```

Testing for uniform DIF on mental health items

Hide

```
mod2 <- DIFtree(Y=MH_sf_items,X=covariate,model="Logistic",type = "udif",alpha=0.05,nperm=
1000,trace=TRUE)
```

```
##
.....

## Split 1; Item 4
##
.....

## Split 2; Item 5
##
.....

## Split 3; Item 6
##
.....
```

Hide

```
mod2
```

```
##
## Item focussed Trees based on the Logistic Regression Approach (uniform DIF):
##
## Call:
## DIFtree.default(Y = MH_sf_items, X = covariate, model = "Logistic",      type = "udi
f", alpha = 0.05, nperm = 1000, trace = TRUE)
##
## Number of persons: 1391
## Number of items: 6
## DIF items: 4 5 6
##
## Overview of executed splits:
##
##   item variable threshold
## 1    4      Sex          0
## 2    5      age         48
## 3    6      age         84
```

Hide

```
summary(mod2)
```

```
##
## Item focussed Trees based on the Logistic Regression Approach (uniform DIF):
##
## Call:
## DIFtree.default(Y = MH_sf_items, X = covariate, model = "Logistic",      type = "uniform", alpha = 0.05, nperm = 1000, trace = TRUE)
##
## -----
##
## Overview:
##
##   item dif   type variables nosplits
## 1     1 no    ---      ---      ---
## 2     2 no    ---      ---      ---
## 3     3 no    ---      ---      ---
## 4     4 yes uniform      Sex        1
## 5     5 yes uniform      age         1
## 6     6 yes uniform      age         1
##
## Total number of Splits: 3
```

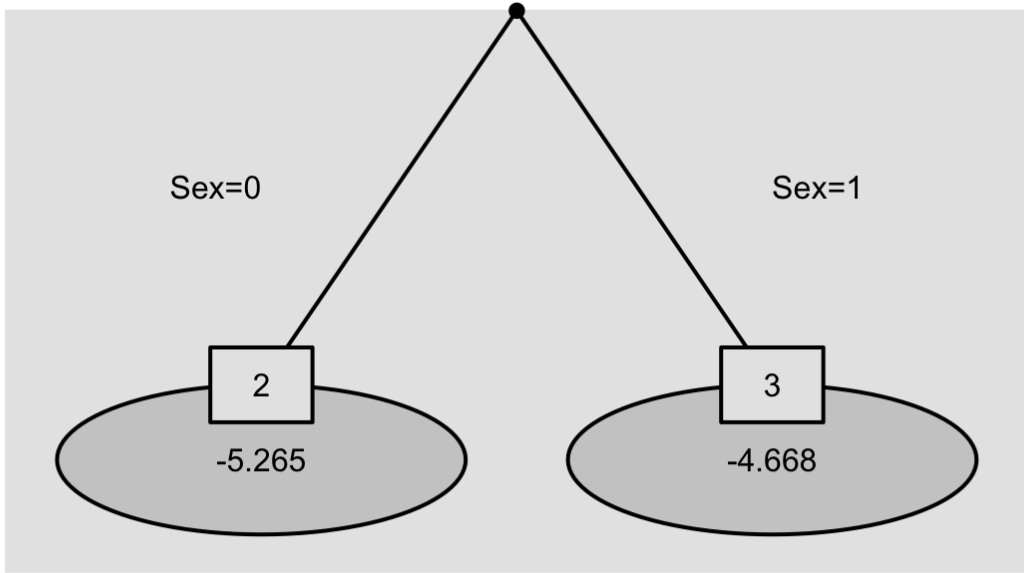
Mental health items 4, 5, and 6 were associated for uniform DIF in item-focussed tree (IFT) model for sex and age.

The tree structure of the IFT model for mental health items

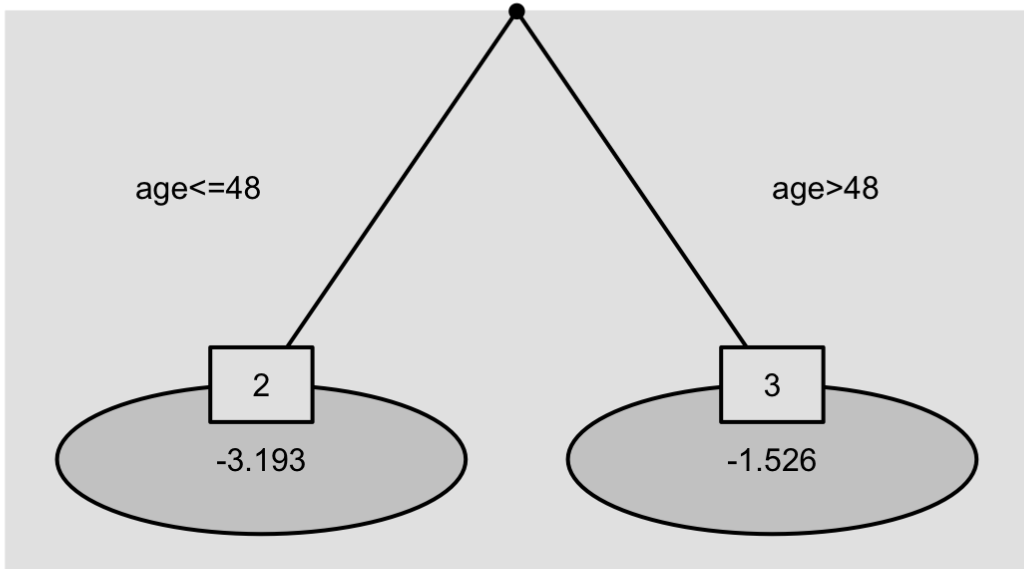
Hide

```
plot(mod2,item=4);plot(mod2,item=5);plot(mod2,item=6)
```

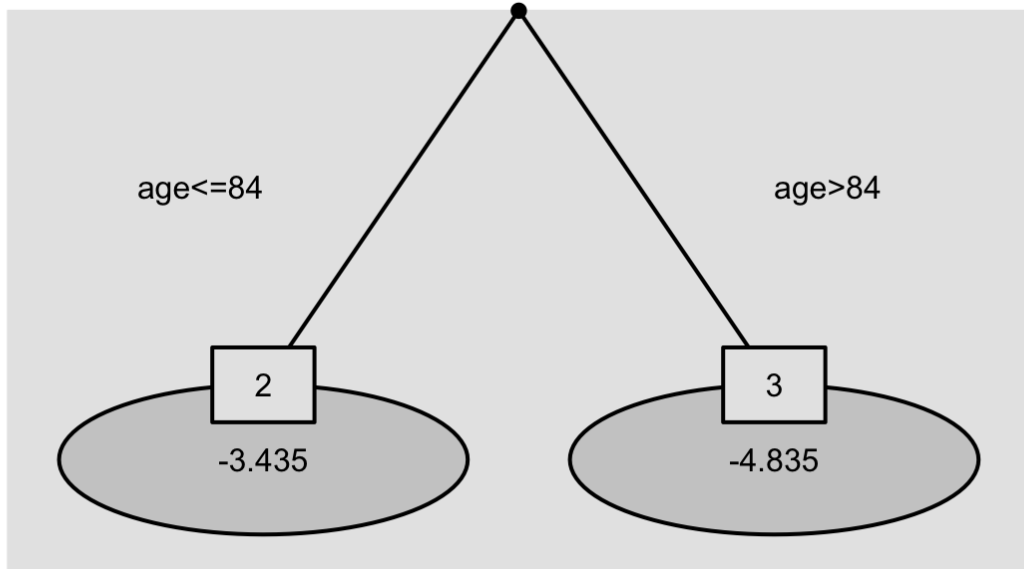

Item 4



Item 5



Item 6



Here, sex=0 denotes female and sex=1 denotes male.

Test statistics and p-values

Hide

```
mod2$devs
```

```
## [1] 17.03099 13.10468 12.93916 10.87760
```

Hide

```
mod2$pvalues
```

```
## [1] 0.000 0.008 0.009 0.028
```

Hide

```
data.frame(
  Item=c(4, 5, 6),
  Variable=c("sex", "age", "age"),
  Test_statistic=c(17.03, "13.10", 12.94),
  p_value=c("<0.00", 0.01, 0.01)
)%>%
datatable(class = 'cell-border stripe', rownames = F, colnames = c("Item", "Variable", "Test
statistic", "p-value"), options = list(pageLength = 3, dom = 't'))
```

Item	Variable	Test statistic	p-value
4	sex	17.03	<0.00
5	age	13.10	0.01
6	age	12.94	0.01

Testing for non-uniform DIF on mental health items

Hide

```
mod4 <- DIFtree(Y=MH_sf_items,X=covariate,model="Logistic",type = "nudif",alpha=0.05,nperm
=1000,trace=TRUE)
```

```
##
.....
```

Hide

```
mod4
```

```
##
## Item focussed Trees based on the Logistic Regression Approach (non-uniform DIF):
##
## Call:
## DIFtree.default(Y = MH_sf_items, X = covariate, model = "Logistic",      type = "nudi
f", alpha = 0.05, nperm = 1000, trace = TRUE)
##
## Number of persons: 1391
## Number of items: 6
## DIF items: no DIF item
##
## Overview of executed splits:
## no split performed
```

Hide

```
summary(mod4)
```



```
##
## Item focussed Trees based on the Logistic Regression Approach (non-uniform DIF):
##
## Call:
## DIFtree.default(Y = MH_sf_items, X = covariate, model = "Logistic", type = "nudi
f", alpha = 0.05, nperm = 1000, trace = TRUE)
##
## -----
##
## Overview:
##
##   item dif type variables nosplits
## 1     1  no  ---         ---      ---
## 2     2  no  ---         ---      ---
## 3     3  no  ---         ---      ---
## 4     4  no  ---         ---      ---
## 5     5  no  ---         ---      ---
## 6     6  no  ---         ---      ---
##
## Total number of Splits:
```

None of the mental health items were associated for non-uniform DIF in IFT model for sex and age.

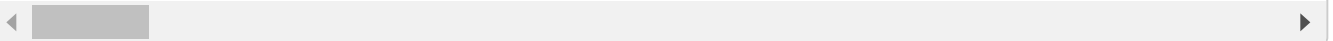
Testing for uniform DIF on physical health items

Hide

```
mod1 <- DIFtree(Y=PH_sf_items,X=covariate,model="Logistic",type = "udif",alpha=0.05,nperm=
1000,trace=TRUE)
```

```
##
.....

## Split 1; Item 3
##
.....
```



Hide

```
mod1
```

```
##
## Item focussed Trees based on the Logistic Regression Approach (uniform DIF):
##
## Call:
## DIFtree.default(Y = PH_sf_items, X = covariate, model = "Logistic", type = "uniform", alpha = 0.05, nperm = 1000, trace = TRUE)
##
## Number of persons: 1391
## Number of items: 6
## DIF items: 3
##
## Overview of executed splits:
##
## item variable threshold
## 1 3 Sex 0
```

Hide

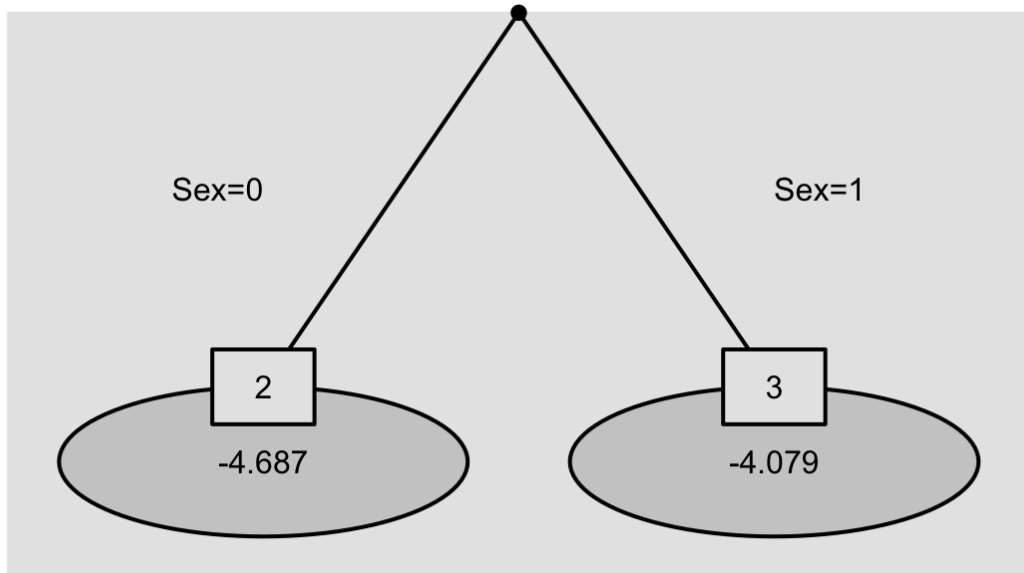
```
summary(mod1)
```

```
##
## Item focussed Trees based on the Logistic Regression Approach (uniform DIF):
##
## Call:
## DIFtree.default(Y = PH_sf_items, X = covariate, model = "Logistic", type = "uniform", alpha = 0.05, nperm = 1000, trace = TRUE)
##
## -----
##
## Overview:
##
## item dif type variables nosplits
## 1 1 no --- --- ---
## 2 2 no --- --- ---
## 3 3 yes uniform Sex 1
## 4 4 no --- --- ---
## 5 5 no --- --- ---
## 6 6 no --- --- ---
##
## Total number of Splits: 1
```

Hide

```
plot(mod1,item=3)
```

Item 3



Here, sex=0 denotes female and sex=1 denotes male.

Test statistics and p-values

Hide

```
mod1$devs
```

```
## [1] 10.277787 8.134947
```

Hide

```
mod1$pvalues
```

```
## [1] 0.002 0.108
```

Hide

```
data.frame(
  Item=c(3),
  Variable=c("sex"),
  Test_statistic=c("10.28"),
  p_value=c("0.01")
)%>%
datatable(class = 'cell-border stripe', rownames = F,colnames = c("Item", "Variable", "Test
t statistic", "p-value"), options = list(pageLength = 2,dom = 't'))
```

Item	Variable	Test statistic	p-value
3	sex	10.28	0.01

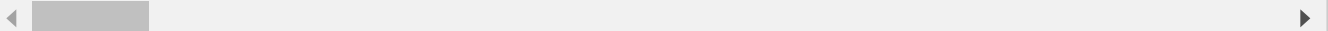
Testing for non-uniform DIF on physical health items

Hide

```
mod3 <- DIFtree(Y=PH_sf_items,X=covariate,model="Logistic",type = "nudif",alpha=0.05,nperm
=1000,trace=TRUE)
```

```
##
```

```
.....
```



Hide

```
mod3
```

```
##
## Item focussed Trees based on the Logistic Regression Approach (non-uniform DIF):
##
## Call:
## DIFtree.default(Y = PH_sf_items, X = covariate, model = "Logistic",      type = "nudi
f", alpha = 0.05, nperm = 1000, trace = TRUE)
##
## Number of persons: 1391
## Number of items: 6
## DIF items: no DIF item
##
## Overview of executed splits:
## no split performed
```

Hide

```
summary(mod3)
```

```

##
## Item focussed Trees based on the Logistic Regression Approach (non-uniform DIF):
##
## Call:
## DIFtree.default(Y = PH_sf_items, X = covariate, model = "Logistic",      type = "nudi
f", alpha = 0.05, nperm = 1000, trace = TRUE)
##
## -----
##
## Overview:
##
##   item dif type variables nosplits
## 1     1  no  ---         ---      ---
## 2     2  no  ---         ---      ---
## 3     3  no  ---         ---      ---
## 4     4  no  ---         ---      ---
## 5     5  no  ---         ---      ---
## 6     6  no  ---         ---      ---
##
## Total number of Splits:

```

None of the physical health items were associated for non-uniform DIF in IFT model for sex and age.