U3Paper_Option_B

youngwoo Kwon

2021 4 21

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(arm)
```

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loading required package: lme4

##
## arm (Version 1.11-2, built: 2020-7-27)

## Working directory is C:/Users/FOS_15/Desktop/WS21/Stats 485/Paper3
```

```
library(ggplot2)
```

Summary : We analysed the F.-S. model. We first reproduced their results with logistic regression using glm() and bayesglm(). Then, we used different variables to find a better model compare to the F.-S.'s model. As a result, we got a better model which showed lower AIC and lower deviance than the original model. We plotted the order of original model and preffered model and get the correlation value. Although two models have overlapping variables, the result states that the order of the models might be distinct from each other.

A : reconstruct measurement models similar to the one used by Finn and Servoss

```
rawdata <- read.csv("http://dept.stat.lsa.umich.edu/~bbh/s485/data/security_wide.csv")
rawdata %>% complete.cases() %>% table()
```

```
## .
## FALSE  TRUE
##    44   612
```

We have to compare the result of the models later. To facilitate comparison (to avoid the absence of index on one side) I will remove missing values in advance.

```
rawdata = rawdata %>% na.omit()
security <- rawdata[c(1,4,5,9,8,15,18,7)]
head(security)
```

```
##   school q38c q38d q38h q38g q38n q39a q38f
## 1 id1011    0    0    1    0    0    1    0
## 2 id1021    1    1    0    1    1    1    0
## 3 id1022    0    0    0    0    0    0    0
## 4 id1031    0    0    1    1    1    1    0
## 5 id1033    0    0    0    0    1    1    0
## 6 id1041    0    0    0    1    1    1    0
```

We don't have to worry about the missing datas now.

```
sum_index <- security %>%
  dplyr::select(-school) %>% rowSums()
table(sum_index)
```

```
## sum_index
##   0   1   2   3   4   5   6
##  81 146 184 119  64  16   2
```

```
security0 <- security %>% gather("item", "response", -school)
head(security0)
```

```
##   school item response
## 1 id1011 q38c        0
## 2 id1021 q38c        1
## 3 id1022 q38c        0
## 4 id1031 q38c        0
## 5 id1033 q38c        0
## 6 id1041 q38c        0
```

```
rasch0 <- glm(response ~ item + school -1, family = "binomial", data = security0)
head(coef(rasch0), 10)
```

```
##     itemq38c     itemq38d     itemq38f     itemq38g     itemq38h     itemq38n
##  -4.24872351  -2.58132594  -0.14751400  -1.44274198  -2.05607672   0.03936898
##     itemq39a schoolid1021 schoolid1022 schoolid1031
##   0.56660141   2.73227774 -17.76083191   1.76881758
```

```
rasch1 <- bayesglm(response ~ item + school - 1, data = security0)
```

```
## Warning in model.matrixBayes(object = mt, data = data, contrasts.arg =
## contrasts, : variable 'item' converted to a factor
```

```
head(coef(rasch1), 10)
```

```
##     itemq38c     itemq38d     itemq38f     itemq38g     itemq38h     itemq38n
##   0.02335348   0.10341088   0.43671255   0.22758172   0.15242565   0.46938940
##     itemq39a schoolid1021 schoolid1022 schoolid1031
##   0.55925101   0.42896082  -0.27934468   0.28724367
```

B : consider whether simple variations of the model specification might improve it

In a new data, I added the question Q38j 'Enforce a strict dress code'. Question 38c and 38d seem similar. So I removed the Q38d from the data. Q38f 'Use one or more random dog sniffs to check for drugs' did not make sense for me. So I removed Q38f.

```
security_new <- rawdata[c(1,4,9,8,15,18,11)]
head(security_new)
```

```
##   school q38c q38h q38g q38n q39a q38j
## 1 id1011    0    1    0    0    1    0
## 2 id1021    1    0    1    1    1    1
## 3 id1022    0    0    0    0    0    0
## 4 id1031    0    1    1    1    1    1
## 5 id1033    0    0    0    1    1    0
## 6 id1041    0    0    1    1    1    1
```

```
security_new0 <- security_new %>% gather("item", "response", -school)
head(security_new0)
```

```
##   school item response
## 1 id1011 q38c        0
## 2 id1021 q38c        1
## 3 id1022 q38c        0
## 4 id1031 q38c        0
## 5 id1033 q38c        0
## 6 id1041 q38c        0
```

```
rasch2 <- glm(response ~ item + school -1, family = "binomial", data = security_new0)
head(coef(rasch2), 10)
```

```
##      itemq38c      itemq38g      itemq38h      itemq38j      itemq38n
## -4.334269e+00 -1.522525e+00 -2.127519e+00  5.124067e-01 -9.160632e-02
##      itemq39a  schoolid1021  schoolid1022  schoolid1031  schoolid1033
##  4.014357e-01  3.669851e+00 -1.796160e+01  3.669851e+00  2.760727e-13
```

```
rasch3 <- bayesglm(response ~ item + school - 1, data = security_new0)
```

```
## Warning in model.matrixBayes(object = mt, data = data, contrasts.arg =
## contrasts, : variable 'item' converted to a factor
```

```
head(coef(rasch3), 10)
```

```
##     itemq38c     itemq38g     itemq38h     itemq38j     itemq38n     itemq39a
##  0.022594494  0.226822488  0.151666517  0.578097352  0.468629760  0.558491173
## schoolid1021 schoolid1022 schoolid1031 schoolid1033
##  0.494106489 -0.331062465  0.494106489 -0.001039664
```

C : compare the alternative Rasch models you will have fit, arriving at a recommended model

```
rasch0$aic
```

```
## [1] 4471.554
```

```
rasch1$aic
```

```
## [1] 4846.643
```

```
rasch2$aic
```

```
## [1] 4245.735
```

```
rasch3$aic
```

```
## [1] 4628.027
```

The model Rasch2 shows the lowest AIC. Rasch0 shows the second-lowest AIC.

```
rasch0$deviance
```

```
## [1] 3235.554
```

```
rasch1$deviance
```

```
## [1] 570.5367
```

```
rasch2$deviance
```

```
## [1] 3011.735
```

```
rasch3$deviance
```

```
## [1] 528.6911
```

Rasch4 has the lowest deviance. Rasch2 has the second-lowest deviance. Comparing the results, it seems the second data set performs the better result. between glm() model and bayesglm() model, I choose the glm() model, the one with lower AIC, since I prefer AIC than deviance when selecting the model.

D : compare Finn and Servoss's school security measurement to your preferred alternative.

```
rasch2_index <- as.vector(c(schoolid1011=0, coef(rasch2)[-(1:6)]))

rasch2_matrix <- matrix(c(as.vector(security$school), rasch2_index), ncol = 2)
head(rasch2_matrix)
```

```
##      [,1]     [,2]
## [1,] "id1011" "0"
## [2,] "id1021" "3.66985115967302"
## [3,] "id1022" "-17.9616040402437"
## [4,] "id1031" "3.66985115967313"
## [5,] "id1033" "2.76072654316916e-13"
## [6,] "id1041" "2.14003632437597"
```

```
rasch2_ordered0 <- rasch2_matrix[order(rasch2_matrix[,2]),]
head(rasch2_ordered0)
```

```
##      [,1]     [,2]
## [1,] "id1781" "-1.14468502860331"
## [2,] "id1661" "-1.14468502860333"
## [3,] "id2081" "-1.14468502860333"
## [4,] "id1702" "-1.14468502860334"
## [5,] "id1732" "-1.14468502860334"
## [6,] "id1791" "-1.14468502860334"
```

This is the order of the school using rasch2, glm() model with security_new data.

```
rasch2_ordered1 = cbind(rasch2_ordered0, c(1:612))
rasch2_ordered2 = rasch2_ordered1[order(rasch2_ordered1[,1]),]
head(rasch2_ordered2)
```

```
##      [,1]     [,2]                   [,3]
## [1,] "id1011" "0"                    "205"
## [2,] "id1021" "3.66985115967302"     "609"
## [3,] "id1022" "-17.9616040402437"    "202"
## [4,] "id1031" "3.66985115967313"     "612"
## [5,] "id1033" "2.76072654316916e-13" "573"
## [6,] "id1041" "2.14003632437597"     "402"
```

The first column of this matrix is school names. The second column is their glm() coefficients. The third column is their order.

Let's repeat this procedure for the F.-S. model, rasch0.

```
rasch0_index <- as.vector(c(schoolid1011=0, coef(rasch0)[-(1:7)]))
rasch0_matrix <- matrix(c(as.vector(security$school), rasch0_index), ncol = 2)
rasch0_ordered0 <- rasch0_matrix[order(rasch0_matrix[,2]),]
rasch0_ordered1 = cbind(rasch0_ordered0, c(1:612))
rasch0_ordered2 = rasch0_ordered1[order(rasch0_ordered1[,1]),]
head(rasch0_ordered2)
```

```
##      [,1]     [,2]                   [,3]
## [1,] "id1011" "0"                    "228"
## [2,] "id1021" "2.73227773914541"     "412"
## [3,] "id1022" "-17.7608319146448"    "153"
## [4,] "id1031" "1.76881757903712"     "363"
## [5,] "id1033" "3.68920363618418e-13" "435"
## [6,] "id1041" "0.892167222603639"    "237"
```

Let's combine the order.

```
rasch_comp <- cbind(Rasch0 = as.numeric(rasch0_ordered2[,3]), Rasch2 = as.numeric(rasch2_ordered2[,3]))
rasch_comp1 <- as.data.frame(rasch_comp[order(rasch_comp[,1]),])
head(rasch_comp)
```

```
##      Rasch0 Rasch2
## [1,]    228    205
## [2,]    412    609
## [3,]    153    202
## [4,]    363    612
## [5,]    435    573
## [6,]    237    402
```

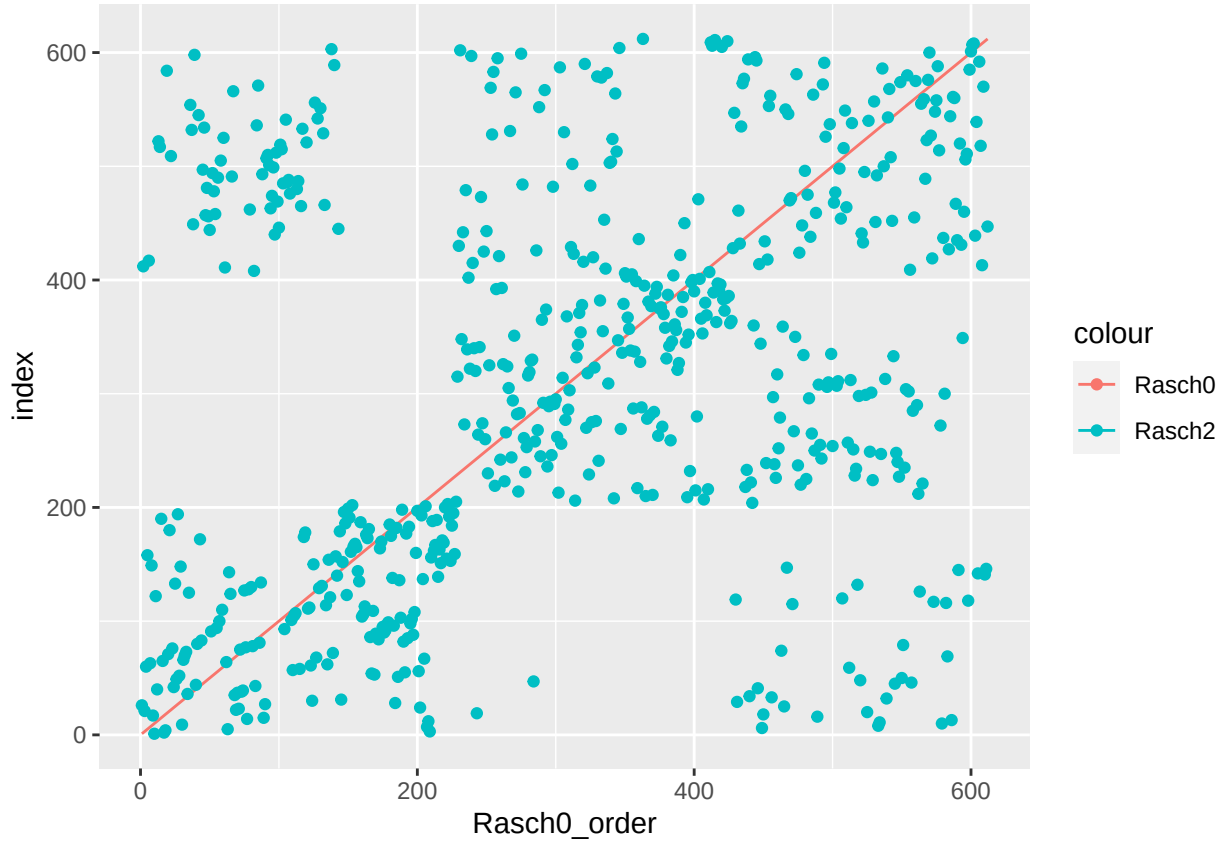
```
head(rasch_comp1)
```

```
##   Rasch0 Rasch2
## 1      1     26
## 2      2    412
```

```
## 3      3     21
## 4      4     60
## 5      5    158
## 6      6    417
```

The first row is index from the F.-S. model and the second row is index from the fixed (preffered) model.

```
ggplot(rasch_comp1, aes(x = Rasch0)) +
geom_line(aes(y = Rasch0, colour = "Rasch0")) +
geom_point(aes(y = Rasch2, colour = "Rasch2")) +
labs(y = "index", x = "Rasch0_order")
```




```
cor(x = rasch_comp1$Rasch0, y = rasch_comp1$Rasch2)
```

```
## [1] 0.3130139
```