Predictive Models in Education

INFO 5200 Learning Analytics: Week 5 Homework

Cole Walsh, 4399966

In this homework, you will first contribute to generating some assessment data that will be used by everyone in the class. Then, you will download the assessment data, score it, and analyze its psychometric properties. It helps to have carefully read *HLA Ch.3, Measurement and its uses in learning analytics.*

Step 1: Take the Assessment

The goal here is to generate some realistic data. You will see 12 questions to answer in 6 minutes (please time yourself). Do not use the internet or other help. Just try your best; you will of course not be graded based on your score on this 6-minute assessment.

Step 2: Download and Import Data

You can open the available assessment responses following this link.

Two things to note:

1. The responses are shown as text. To use psychometric methods you need to convert this to 0/1 data i.e. score the responses.
2. The first answer row has a score of 12/12. This means that those text responses are all correct. You can use this fact for scoring.

Download the responses as a CSV: File > Download as > Comma-separated values.

Import the CSV file into R/Rstudio using the `read.csv()` function.

```r
# load the csv here
dat = read.csv('INFO5200_HW4_Responses.csv')
```

Step 3: Score Assessment Data

The goal is to create a new dataset that is 12 columns wide and N columns long (N = num. responses) with a value of 1 for correct answers, 0 for incorrect answers. Complete the following steps:

- First, make sure that the data you imported looks correct e.g. using `head()`.
- Second, remove *Timestamp*, *Score*, and *Enter.your.Cornell.NetID* columns from the dataset. You don't need them anymore.
- Third, save the first row (the one with the 12/12 score) as the answer key into a new variable *corres*.
- Fourth, remove the row from the dataset, it's not real data.
- Fifth, to score the responses, create a dataset that repeats the correct answers (*corres*) as many times as there are responses; then compare the real responses to the new matrix of correct responses. This isn't the best way and I provide you some code below. Feel free to use another way to get to the solution. Be sure to convert the TRUE/FALSE responses into 1/0, e.g. using *1 * (response_data == corres_mat)* where == does the scoring and 1*(. . .) takes care of the conversion to binary.

```r
# first
head(dat)
```

```
##            Timestamp   Score
## 1 2/18/2019 23:44:50 12 / 12
## 2  2/19/2019 8:06:02  7 / 12
## 3  2/19/2019 8:27:22  9 / 12
## 4  2/19/2019 8:35:29  7 / 12
## 5 2/19/2019 10:00:50  5 / 12
## 6 2/19/2019 10:15:52  8 / 12
##                                              Radiocarbon.is.produced.in.the.atmosphere.as.a.result.of
## 1                   A. collision between fast neutrons and nitrogen nuclei present in the atmosphere
## 2                                  B. action of ultraviolet light from the sun on atmospheric oxygen
## 3 C. action of solar radiations particularly cosmic rays on carbon dioxide present in the atmosphere
## 4                   A. collision between fast neutrons and nitrogen nuclei present in the atmosphere
## 5 C. action of solar radiations particularly cosmic rays on carbon dioxide present in the atmosphere
## 6                   A. collision between fast neutrons and nitrogen nuclei present in the atmosphere
##   It.is.easier.to.roll.a.stone.up.a.sloping.road.than.to.lift.it.vertical.upwards.because
## 1                              D. work done in rolling a stone is less than in lifting it
## 2                              D. work done in rolling a stone is less than in lifting it
## 3              C. work done in both is same but the rate of doing work is less in rolling
## 4              C. work done in both is same but the rate of doing work is less in rolling
## 5                              D. work done in rolling a stone is less than in lifting it
## 6              C. work done in both is same but the rate of doing work is less in rolling
##   The.absorption.of.ink.by.blotting.paper.involves
## 1                   B. capillary action phenomenon
## 2         C. diffusion of ink through the blotting
## 3                   B. capillary action phenomenon
## 4         C. diffusion of ink through the blotting
## 5         C. diffusion of ink through the blotting
## 6         C. diffusion of ink through the blotting
##                                           A.siphon.will.fail.to.work.if
## 1 B. the level of the liquid in the two vessels are at the same height
## 2          A. the densities of the liquid in the two vessels are equal
## 3 B. the level of the liquid in the two vessels are at the same height
## 4          A. the densities of the liquid in the two vessels are equal
## 5          A. the densities of the liquid in the two vessels are equal
## 6          A. the densities of the liquid in the two vessels are equal
##   Large.transformers..when.used.for.some.time..become.very.hot.and.are.cooled.by.circulating.oil..Th
## 1                                                                                C. both the heating
## 2                                                                                                  A
## 3                                                                                C. both the heating
## 4                                                                                C. both the heating
## 5                                                                                C. both the heating
## 6                                                                                C. both the heating
##   Nuclear.sizes.are.expressed.in.a.unit.named Light.year.is.a.unit.of
## 1                                    A. Fermi             B. distance
## 2                                    A. Fermi             B. distance
## 3                                    A. Fermi             B. distance
## 4                                 B. angstrom             B. distance
## 5                                 B. angstrom             B. distance
## 6                                    A. Fermi                 A. time
##                                           Mirage.is.due.to
## 1 A. unequal heating of different parts of the atmosphere
## 2 A. unequal heating of different parts of the atmosphere
## 3 A. unequal heating of different parts of the atmosphere
## 4 A. unequal heating of different parts of the atmosphere
## 5 A. unequal heating of different parts of the atmosphere
## 6 A. unequal heating of different parts of the atmosphere
##   Light.from.the.Sun.reaches.us.in.nearly
## 1                            C. 8 minutes
## 2                            C. 8 minutes
## 3                            C. 8 minutes
## 4                            C. 8 minutes
## 5                            A. 2 minutes
## 6                            C. 8 minutes
##   Stars.appears.to.move.from.east.to.west.because
## 1          B. the earth rotates from west to east
## 2          B. the earth rotates from west to east
## 3          B. the earth rotates from west to east
## 4          B. the earth rotates from west to east
## 5          C. the earth rotates from east to west
## 6          B. the earth rotates from west to east
##   Pa.Pascal..is.the.unit.for
## 1                B. pressure
## 2                B. pressure
## 3            D. conductivity
## 4                B. pressure
## 5                B. pressure
## 6                B. pressure
##
## 1 D. they are nearer to earth and hence we receive a greater amount of light and, therefore minor va
## 2
## 3 D. they are nearer to earth and hence we receive a greater amount of light and, therefore minor va
## 4
## 5
## 6 D. they are nearer to earth and hence we receive a greater amount of light and, therefore minor va
##   Enter.your.Cornell.NetID
## 1
## 2                    zad22
## 3                   cds238
## 4                   cjw295
## 5                   jm2682
## 6                     akx2
```

```r
# second
dat2 <- dat[, -c(1, 2, 15)]

# third
corres <- dat2[1, ]

# fourth
response_data <- dat2[2:nrow(dat2), ]

# fifth
corres_mat = corres[rep(1, nrow(response_data)),] # dat is the response dataset
response_data <- 1 * (response_data == corres_mat)
```

```
# Quick fix: cut long column names for prettier output:
colnames(response_data) = substr(colnames(response_data), 0, 12)
head(response_data)
```

```
##   Radiocarbon. It.is.easier The.absorpti A.siphon.wil Large.transf
## 2            0            1            0            0            0
## 3            0            0            1            1            1
## 4            1            0            0            0            1
## 5            0            1            0            0            1
## 6            1            0            0            0            1
## 7            1            1            0            1            1
##   Nuclear.size Light.year.i Mirage.is.du Light.from.t Stars.appear
## 2            1            1            1            1            1
## 3            1            1            1            1            1
## 4            0            1            1            1            1
## 5            0            1            1            0            0
## 6            1            0            1            1            1
## 7            0            1            1            1            1
##   Pa.Pascal..i Planets.do.n
## 2            1            0
## 3            0            1
## 4            1            0
## 5            1            0
## 6            1            1
## 7            1            1
```

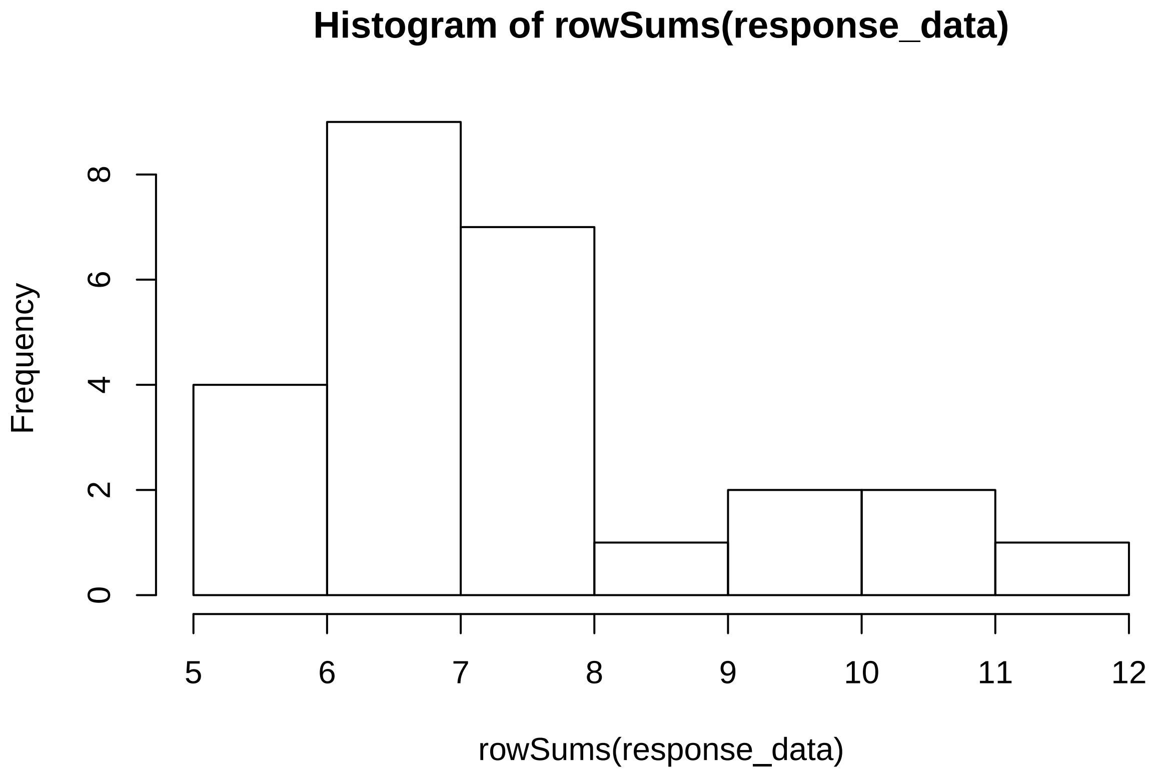
Step 4: Basic Psychometrics

Resource: Wikipedia on Cronbach's alpha

The goal here is to get an overview of the assessment and item characteristics.

1. Compute the total score for each person (i.e. row) and plot the distribution in a histogram.

```
hist(rowSums(response_data))
```



2. Compute how difficult the items were in terms of the proportion of respondents who got them wrong. Which was the hardest/easiest item?

```
colMeans(response_data)
```

```
## Radiocarbon. It.is.easier The.absorpti A.siphon.wil Large.transf
##    0.3076923    0.4615385    0.3846154    0.5384615    0.8461538
## Nuclear.size Light.year.i Mirage.is.du Light.from.t Stars.appear
##    0.5000000    0.9230769    0.9230769    0.6923077    0.8846154
## Pa.Pascal..i Planets.do.n
##    0.9230769    0.4615385
```

The easiest items were questions 7, 8, and 11 where 24 out of 26 people got the question correct. The hardest item was the first question where only 8 people out of 26 got the question correct.

3. Assess the reliability of the assessment using Cronbach's alpha and how the reliability would change if an item were dropped from the assessment. Use the *alpha()* function from the *psych* package to compute these. **Based on the output, what is the (raw) alpha coefficient and is there an item that could be dropped to increase the reliability of the assessment?**

```
# install.packages("psych")
library(psych)
alpha(response_data)
```

```
## Some items ( Radiocarbon. It.is.easier Large.transf Light.year.i Pa.Pascal..i ) were negatively corr
## probably should be reversed.
## To do this, run the function again with the 'check.keys=TRUE' option
```

```
##
## Reliability analysis
## Call: alpha(x = response_data)
##
##   raw_alpha std.alpha G6(smc) average_r S/N  ase mean   sd median_r
##       0.32      0.23    0.51     0.025 0.3 0.19 0.65 0.15    -0.01
##
##  lower alpha upper     95% confidence boundaries
## -0.05 0.32 0.69
##
##  Reliability if an item is dropped:
##              raw_alpha std.alpha G6(smc) average_r   S/N alpha se var.r
## Radiocarbon.      0.35     0.250    0.48    0.0293 0.332     0.18 0.035
## It.is.easier      0.30     0.179    0.47    0.0194 0.218     0.19 0.037
## The.absorpti      0.34     0.249    0.51    0.0292 0.331     0.18 0.032
## A.siphon.wil      0.18     0.098    0.39    0.0098 0.109     0.23 0.033
## Large.transf      0.38     0.322    0.51    0.0415 0.476     0.17 0.035
## Nuclear.size      0.27     0.168    0.45    0.0180 0.201     0.20 0.036
## Light.year.i      0.31     0.239    0.51    0.0278 0.314     0.19 0.037
## Mirage.is.du      0.33     0.257    0.48    0.0304 0.345     0.19 0.034
## Light.from.t      0.27     0.205    0.44    0.0229 0.257     0.20 0.035
## Stars.appear      0.27     0.154    0.46    0.0163 0.182     0.20 0.034
## Pa.Pascal..i      0.37     0.334    0.54    0.0437 0.503     0.18 0.033
## Planets.do.n      0.16     0.069    0.40    0.0067 0.075     0.24 0.034
##                 med.r
## Radiocarbon.  0.0e+00
## It.is.easier -2.0e-02
## The.absorpti -2.0e-02
## A.siphon.wil -2.2e-02
## Large.transf  1.3e-02
## Nuclear.size -2.2e-02
## Light.year.i  0.0e+00
## Mirage.is.du  7.8e-21
## Light.from.t -2.2e-02
## Stars.appear -2.2e-02
## Pa.Pascal..i  7.8e-21
## Planets.do.n -2.0e-02
##
##  Item statistics
##               n  raw.r std.r  r.cor r.drop mean   sd
## Radiocarbon. 26  0.253 0.258  0.158 -0.015 0.31 0.47
## It.is.easier 26  0.396 0.398  0.297  0.115 0.46 0.51
## The.absorpti 26  0.300 0.260  0.127  0.018 0.38 0.50
## A.siphon.wil 26  0.588 0.534  0.552  0.347 0.54 0.51
## Large.transf 26  0.086 0.088 -0.065 -0.123 0.85 0.37
## Nuclear.size 26  0.446 0.418  0.351  0.172 0.50 0.51
## Light.year.i 26  0.225 0.280  0.114  0.072 0.92 0.27
## Mirage.is.du 26  0.142 0.243  0.139 -0.013 0.92 0.27
## Light.from.t 26  0.424 0.350  0.301  0.170 0.69 0.47
## Stars.appear 26  0.386 0.442  0.373  0.213 0.88 0.33
## Pa.Pascal..i 26 -0.026 0.056 -0.132 -0.177 0.92 0.27
## Planets.do.n 26  0.619 0.577  0.578  0.388 0.46 0.51
##
## Non missing response frequency for each item
##                 0    1 miss
## Radiocarbon. 0.69 0.31    0
## It.is.easier 0.54 0.46    0
## The.absorpti 0.62 0.38    0
## A.siphon.wil 0.46 0.54    0
## Large.transf 0.15 0.85    0
## Nuclear.size 0.50 0.50    0
## Light.year.i 0.08 0.92    0
## Mirage.is.du 0.08 0.92    0
## Light.from.t 0.31 0.69    0
## Stars.appear 0.12 0.88    0
## Pa.Pascal..i 0.08 0.92    0
## Planets.do.n 0.54 0.46    0
```

The raw alpha coefficient is 0.32 (95% confidence interval stretched 0.37 in either direction). Dropping either question 5 or 11 would result in an increase in alpha of 0.05, so either of these questions could be dropped, but the reliability of the asssessment is still very low.

4. Conduct an exploratory factor analysis (EFA) of the assessment responses. First, you need to find the optimal number of factors. To do this, use the *fa.parallel()* function from the *psych* package, passing just the dataset as the input. The resulting Scree Plot and the function output below will help you choose the number of factors. Next, you run a factor analysis model using the *fa()* function specifying the *nfactors =* parameter with the optimal number of factors. You then take the result of running *fa()* and plot a diagram of the result using the *fa.diagram()* function which just takes the output of *fa()* as its input. Describe your observation in 1-2 sentences.

```r
fa.parallel(response_data)
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : A loading greater than abs(1) was detected. Examine the loadings
## carefully.
```

```
## The estimated weights for the factor scores are probably incorrect.  Try a different factor extracti
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : An ultra-Heywood case was detected. Examine the results carefully
```

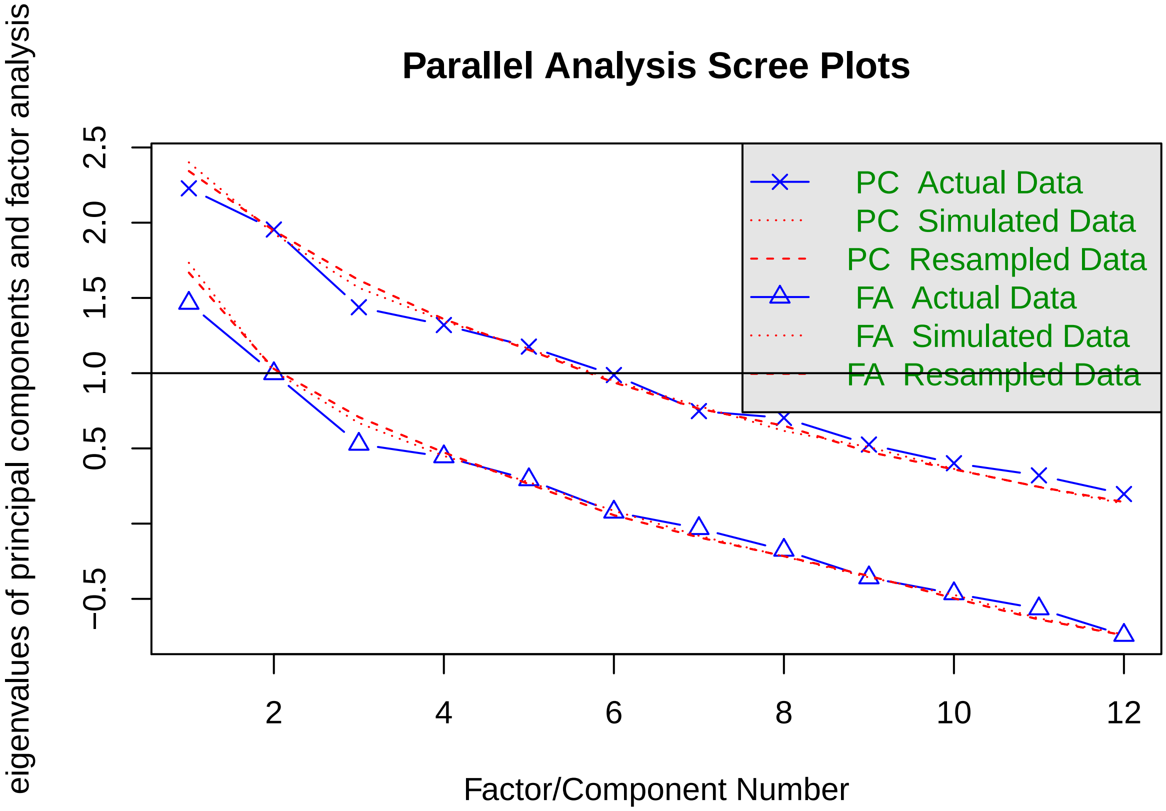
```
## Warning in cor.smooth(R): Matrix was not positive definite, smoothing was
## done
```

```
## Warning in cor.smooth(R): Matrix was not positive definite, smoothing was
## done
```

```
## Warning in cor.smooth(r): Matrix was not positive definite, smoothing was
## done
```

```
## In factor.scores, the correlation matrix is singular, an approximation is used
```

```
## Warning in cor.smooth(r): Matrix was not positive definite, smoothing was
## done
## Warning in cor(sampledata, use = use): the standard deviation is zero

## Warning in cor(sampledata, use = use): the standard deviation is zero

## Warning in cor(sampledata, use = use): the standard deviation is zero

## Warning in cor(sampledata, use = use): the standard deviation is zero

## Warning in cor(sampledata, use = use): the standard deviation is zero

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : A loading greater than abs(1) was detected. Examine the loadings
## carefully.

## The estimated weights for the factor scores are probably incorrect.  Try a different factor extractio

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : An ultra-Heywood case was detected. Examine the results carefully

## Warning in cor(sampledata, use = use): the standard deviation is zero

## The estimated weights for the factor scores are probably incorrect.  Try a different factor extractio

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : A loading greater than abs(1) was detected. Examine the loadings
## carefully.

## The estimated weights for the factor scores are probably incorrect.  Try a different factor extractio

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : An ultra-Heywood case was detected. Examine the results carefully

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : the standard deviation is zero

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : the standard deviation is zero

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : the standard deviation is zero

## The estimated weights for the factor scores are probably incorrect.  Try a different factor extractio
```

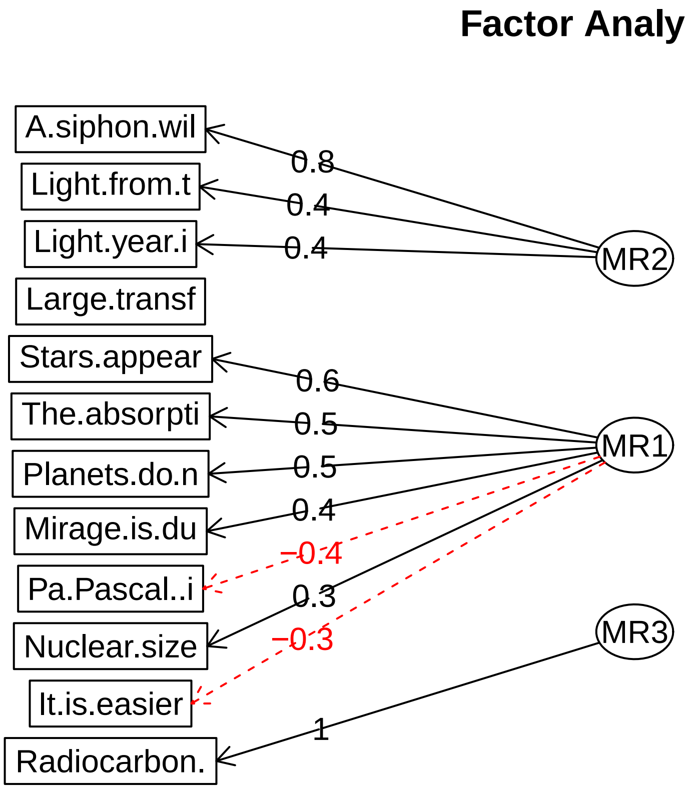


```
## Parallel analysis suggests that the number of factors =  0  and the number of components =  0
```

```
mod <- fa(response_data, nfactors = 3)
```

```
## Loading required namespace: GPArotation
## The estimated weights for the factor scores are probably incorrect.  Try a different factor extracti
```

```
fa.diagram(mod)
```



There is an elbow in the scree plot at nfactors = 3, hence I chose to use 3 factors here. There appears to be a large factor made up of 7 items and a smaller one made of 3 items. One item loaded onto its own factor and one item did not load onto any factors, which I'm not sure I understand.

Step 5: Fit an IRT Model

Resources

- R-blogger article on IRT
- Overview of IRT methods by Columbia School of Public Health

Now you will fit a latent trait model using the IRT approach to estimate the item difficulty (how hard it is) and item discrimination (how well it discriminates between low/high ability students). To do this, install and load the *ltm* package. Then, use the *ltm()* function to fit the model. This function takes a formula as input that specifies how many latent traits to estimate. You just want to estimate one here. The syntax for this formula is *yourDataSet ~ z1* where z1 stands for the first latent trait, which is the student ability. Check the resulting output for Item Difficulty (Dffclt) and Discrimination (Dscrmn). Good items have moderate difficulty and high discrimination. When the difficulty is very high/low, it means everyone/no one gets it right–that's bad. If the discrimination is low, it means that the item is ineffective at distinguishing high from low ability students (they answer it correctly at similar rates). Which items are good/bad?

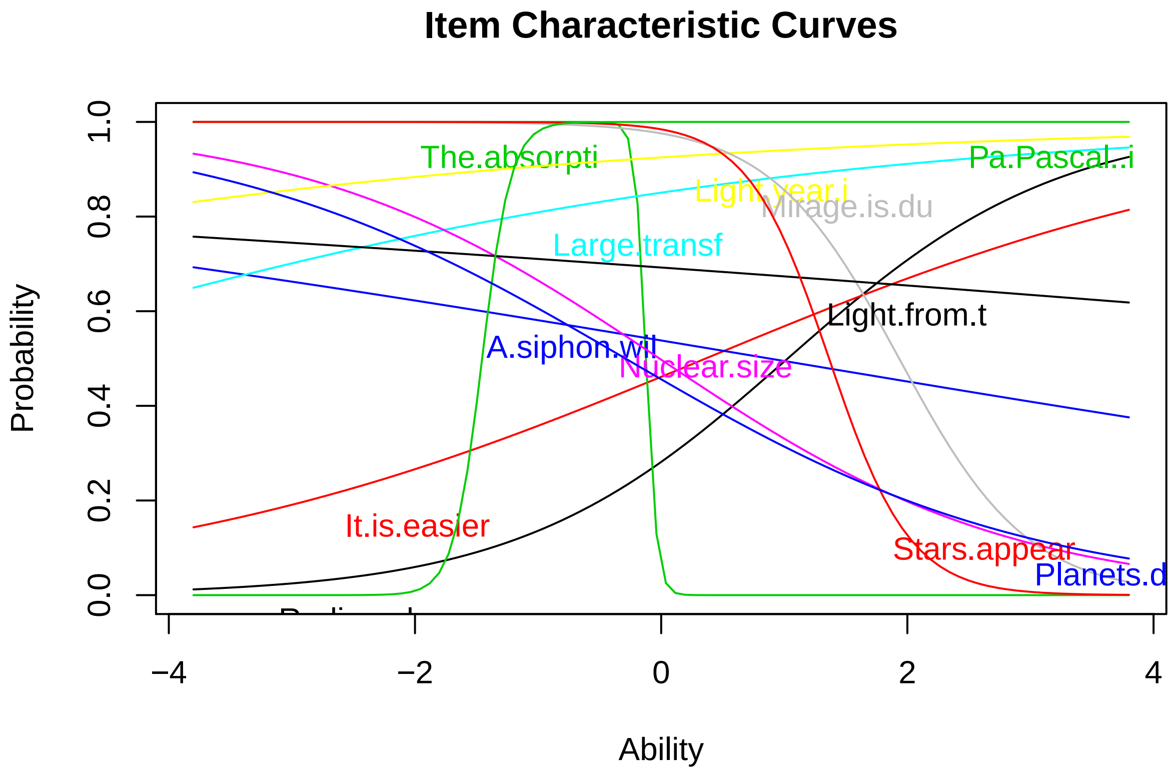
```r
# install.packages("ltm")
library(ltm)

irt.out = ltm(response_data ~ z1)
irt.out
```

```
##
## Call:
## ltm(formula = response_data ~ z1)
##
## Coefficients:
##                 Dffclt   Dscrmn
## Radiocarbon.     1.028    0.912
## It.is.easier     0.361    0.430
## The.absorpti    -0.123  -22.634
## A.siphon.wil     0.879   -0.174
## Large.transf    -5.888    0.295
## Nuclear.size    -0.014   -0.695
## Light.year.i   -10.355    0.243
## Mirage.is.du     1.933   -1.910
## Light.from.t     9.384   -0.086
## Stars.appear     1.364   -3.038
## Pa.Pascal..i    -1.455    8.626
## Planets.do.n    -0.290   -0.606
##
## Log.Lik: -153.665
```

Question 11 is a little easy, but it is very highly discriminating making it a good question. Question 1 is a more difficult, but has the next largest discrimination, which is relatively moderate, so it is also a good question. The only other 'good' question looks to be question 2 since it is slighly difficult and has a little discrimination. All of the other questions suffere from low or negative discrimination (where higher ability students are pedicted to do worse). Question 3, 8 and 10 are particularly bad with large negative discriminations. Questions 4, 5, 6, and 12 also have negative discriminations. These are all bad questions.

Finally, plot the Item Characteristic Curves by passing the output from *irt.out = ltm(. . .)* into the *plot(irt.out)* function. The x-axis is student ability; the y-axis is probability of answering correctly (i.e. "facility" = 1 - difficulty).

```r
plot(irt.out)
```



Optional: If you'd like to explore different IRT models and datasets a bit more to build your intuition, check out this interactive application. Simply uncomment the three lines of code below—but keep them commented out for the submission (i.e. keep eval=F).

```
#install.packages("IRTShiny")
library(IRTShiny)
startIRT()
```

Self-reflection

Briefly summarize your experience on this homework. What was easy, what was hard, what did you learn?

I really enjoyed this homework. I thought the homework was easy, but I've had plenty of experience with R in research and some experience with assessment development, so I don't object to the difficulty of the homework. This was useful to me because I got to think more about IRT, which is something I haven't used so far in my research, but now that I have a better handle on it, I plan on doing just that.

Submit Homework

This is the end of the homework. Please **Knit a PDF report** that shows both the R code and R output and upload it on the EdX platform. Alternatively, you can Knit it as a "doc", open it in Word, and save that as a PDF.

Important: Be sure that all your code is visible. If the line is too long, it gets cut off. If that happens, organize your code on several lines.