Tutorial 1 Factor Analysis

1. Find and load the `bfi` dataset (in the psych package).
   `library(psych) # bfi is autoloaded`
   - what columns contain the Big-Five Inventory data?
   `1:25 called A1-5 C1-5 etc.`
2. Find a package in R that does parallel analysis
   - What is it's name?
   `paran`
   - What is the name of the function?
   `paran`
   - *tutor note*: Share the package and function we wish to work with to the group.
3. Read the help
   - What parameters does this parallel analysis function take?
   - What do they do?
4. Use the function to determine how many factors are in the bfi dataset
   - Do rows with missing data break paran?
   `Yes`
   - Does parallel analysis function need to be given *just* the columns you need to analyse?
   `Yes`
   - How many complete.cases exist in these personality data?
   `sum(complete.cases(bfi)) = 2236`
   `df = bfi[complete.cases(bfi), 1:25]`
   - Run the function?
   `paran::paran(df)`
   - How many factors exist in these personality data?
   `5`
   - What is a scree plot and how do you plot it with this function?
   `paran::paran(df, graph= TRUE)`
5. Find R's built in factor analysis function
   - What is it?
   `factanal`
   - *tutor note* - share this answer with the class if they don't get it
   - What parameters does this function need?
   `data and factors at a minimum`
   - What are its options? Discuss.
6. Run an fa, extracting the predicted number of factors from paran
   `fa = factanal(df, factors = 5)`
   - What does uniqueness mean?
   `how much of each item is NOT explained by the factors`
   - Are items fairly unique in general?

```
yes, often well-over half their variance
```
  ○ Was what you ran by default oblique or orthogonal?
```
orthogonal
```
  ○ What is the name of an oblique rotation?
```
let me google that for you
```
```
promax
```
  ○ *tutor-note* share the correct answer before continuing.
7. Use the oblique rotation
```
fa = factanal(df, factors = 5, rotation= "promax")
```
  ○ Is the structure "simple" now?
```
yes
```
  ○ What does that mean?
```
items have one high loading and all the rest are small
```
  ○ What are the factors? (Name them based on high loadings)
```
look at high-loadings on the print out
```
  ○ What do the empty cells mean?
```
print out hides small values
```
8. Try and alter how the result prints out
  ○ The factor analysis object has a special print method, which supports sorting and hiding small values!
```
print(fa, cutoff = .3, sort = TRUE)
```
  ○ Are the factors independent?
  ○ What component of the print out tells us this?
9. Create scores for each subject (hint, the factor analysis function has a scores parameter)
```
fa = factanal(df, factors = 5, scores = "Bartlett", data = df,
na.action = na.exclude)
```
10. Add these to the dataset.
```
df$f1= fa$scores[,"Factor1"]
```

Bravo!

Extra credit if you finish early

1. Try doing all of this with IQ data set Holzinger from `psych`
2. Do an FA on some of your own data, or… anything else: practise creates skill.
3. Play with the options to paran and factanal

To prepare for next week's tutorials and lectures

1. Install the package `umx`
2. Read the `?umxRAM` help, and run one model from its help examples
3. Advanced credit: Try and re-run one of the factor analyses using `umxFactanal`

Scientific as opposed to statistical Questions:

1. Do you think personality has 5 or 6 major domains?
2. Is the BFI data good?
3. What would happen to the parallel analysis if we sampled facets better?
4. What could go wrong if the data have a hierarchical structure like we know personality does?