Lab 6

Multivariate Statistics with R

This week's lab is the first in a five part block on latent variable models (factor analysis, path analysis, and structural equation modeling). For lots of useful info and links, visit Tim's Multivariate Stats Course page. And now, let's get going!

Task 1: Find and load the `bfi` dataset (in the `psych` package).

- **Question 1.1:** what columns contain the Big-Five Inventory data?

Task 2: Find a package in `R` that does parallel analysis.

- **Question 2.1:** What is it's name?
- **Question 2.2:** What is the name of the function?

Task 3: Read the function documentation (help file).

- **Question 3.1:** What parameters does this parallel analysis function take?
- **Question 3.2:** What do they do?

Task 4: Use the function to determine how many factors are in the `bfi` dataset.

- **Question 4.1:** Assuming that didn't work, what went wrong?
- **Question 4.2:** Does the parallel analysis function need to be given just the columns you need to analyse?
- **Question 4.3:** How many complete cases exist in these personality data?

Task 5: Run the function on the appropriate subset of `bfi`.

- **Question 5.1:** How many factors exist in these personality data?
- **Question 5.2:** What is a scree plot and how do you plot it with this function?

Task 6: Find `R`'s built in factor analysis function.

- **Question 6.1:** Which one is it?
- **Question 6.2:** What parameters does this function need?
- **Question 6.3:** What are its options? Discuss.

Task 7: Run an fa, extracting the predicted number of factors from `paran()`.

- **Question 7.1:** What does uniqueness mean?
- **Question 7.2:** Are items fairly unique in general?
- **Question 7.3:** Was what you ran by default oblique or orthogonal?
- **Question 7.4:** What is the name of an oblique rotation?

Task 8: Use the oblique rotation available in `factanal()`.

- **Question 8.1:** Is the structure "simple" now?
- **Question 8.2:** What does that mean?
- **Question 8.3:** What are the factors? (Name them based on high loading items)
- **Question 8.4:** What do the empty cells mean?

Task 9: Try and alter how the result prints out. Let's say we want to see only loadings > .3 and we want the items sorted by factors that load on them.

Hint: Look for the `print` method in the help file for `factanal()`.

- **Question 9.1:** Are the factors independent?
- **Question 9.2:** What component of the printout tells us this?

Task 10: Create scores for each subject

Hint: The factor analysis function has a `scores` parameter.

Task 11: Add these to the dataset.

Hint: Check the function documentation to find out where the scores are stored.

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?