Psychometrics Project

1. Obtain your dataset

https://tinyurl.com/ds5740-project2-dataset

Make sure to report your dataset in your project report.

2. Instructions

A colleague walks up to you and tells you they've collected some survey data from two different countries. They tell you they think they've developed an excellent Extraversion scale. They are wondering whether they've developed a good scale. You tell them that luckily you've taken a course that introduced you to some statistical tools to evaluate their survey. They're stoked and tell you more about their dataset.

They say that they developed the scale with 6 dimensions: activity level, assertiveness, cheerfulness, excitement-seeking, friendliness, and gregariousness. For each dimension, they developed 4 items. They had participants rate each item from 1-5 from "very inaccurate" to "very accurate" but they *reverse-coded* items that were inversely related to Extraversion. The item "avoid crowds," for example, that was originally scored "1" was *reverse* scored to be "5". They did this reverse coding to try to ensure all relationships between variables would be *positive*.

They mention to you that they have a few questions regarding their dataset:

1. Do the number of dimensions match the theoretical dimensions? What about the items placed into those dimensions – that is, how well do the item placements in the theoretical dimensions match the item placement into the empirical dimensions? (*2 points*)

2. How stable is the estimated structure? Can the structure be generalized to other datasets? Are there any problematic items? (3 *points*)

3. Even if there are no problematic items, are potential issues arising from multidimensionality? What about the assumption of local independence – do any pairs of variables violate this assumption? (*2 points*)

After some thought, they remember that the data are from two countries and wondered about a few more things:

4. Do the empirical dimensions for each country match the theoretical? How well do they match? What about the similarity between both countries' dimensions? (*2 points*)

5. Are the structures for each country stable? What about problematic items? Is local independence violated? (*3 points*)

Finally, they say they have two important questions that they need to report to their boss. First, they want to develop an experiment to tap into Extraversion similarities or differences between the two countries. They wonder if you might have some input based on your results.

6. Do the empirical structures of the countries differ statistically from one another? If there are items that differ, what are they? Speculate on the mechanisms behind the differences or lack of differences between the groups and think about how these differences might be evaluated in an experiment (we haven't covered experimental design yet but speculate about a task that

these two groups could perform that might give you a *behavioral* measure of this difference). (4 *points*)

Second, they are developing an ad campaign that is tailored to different generations. They think that millennials might be more extraverted than boomers but they aren't sure if this is true in each country. They would like training and testing data to know how well they can expect the results to generalize.

7. Can age be predicted by the dimensions of Extraversion for each country (treat age as continuous)? Do the dimensions that predict age differ between the countries? How well can the training model be expected to generalize (report correlation between the predicted and observed results of the test data) (*4 points*)

/*20 points*