For this assignment, we will use data from the Levenson Self-Report Psychopathy Scale (LSRP). This instrument is supposed to be a screening test for Psychopathy, but the underlying dimensionality of the data (the number of factors) is in dispute. Your task is to analyze the data in this regard.

First, let's discuss the LRSP. It consists of 26 statements. People are asked to indicate the degree to which they agree or disagree with each statement, indicated by a number (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree).

The 26 statements are, in order:

1 Success is based on survival of the fittest; I am not concerned about the losers.

2 I find myself in the same kinds of trouble, time after time.

3 For me, what's right is whatever I can get away with.

4 I am often bored.

5 In today's world, I feel justified in doing anything I can get away with to succeed.

6 I find that I am able to pursue one goal for a long time.

7 My main purpose in life is getting as many goodies as I can.

8 I don't plan anything very far in advance.

9 Making a lot of money is my most important goal.

10 I quickly lose interest in tasks I start.

11 I let others worry about higher values; my main concern is with the bottom line.

12 Most of my problems are due to the fact that other people just don't understand me.

13 People who are stupid enough to get ripped off usually deserve it.

14 Before I do anything, I carefully consider the possible consequences.

15 Looking out for myself is my top priority.

16 I have been in a lot of shouting matches with other people.

17 I tell other people what they want to hear so that they will do what I want them to do.

18 When I get frustrated, I often "let off steam" by blowing my top.

19 I would be upset if my success came at someone else's expense.

20 Love is overrated.

21 I often admire a really clever scam.

22 I make a point of trying not to hurt others in pursuit of my goals.

23 I enjoy manipulating other people's feelings.

24 I feel bad if my words or actions cause someone else to feel emotional pain.

25 Even if I were trying very hard to sell something, I wouldn't lie about it.

26 Cheating is not justified because it is unfair to others.

In general, the higher the score is, the higher the likelihood that someone qualifies as a psychopath. To avoid response biases (some people say yes to everything), some of the items are "inversely coded" – e.g. item 14, the higher the response on that, the less likely someone is to be a psychopath. Your analysis needs to take that into account.

Basic assignment

Your program should be able to do the following – assume the data file is in the same directory as the program file.

1. Load the data from LRSP.xlsx. The first 26 columns contain the data from the 26 statements – in order – above. Rows are individual participants.
2. Reverse the coding of the items that are inversely scored so that a higher score indicates a higher degree of psychopathy. For instance, if someone gives a 5 on item 14, it is really a 1. If someone gives a 2, it is really a 4, whereas a 3 stays a 3. Overall, the following items are inversely coded: 6, 14, 19, 22, 24, 25, 26
3. Remove missing data – in this case, it is probably best to simply delete participants who didn't answer all the questions.
4. Make a figure with a histogram of the distribution of total scores.
5. Do a principal component analysis of the data.
6. Determine how many factors there are – 1, 2, 3, 4 or more?
7. Make a figure that plots individuals in the factor space that resulted from your PCA. Apply suitable labels to the axes (how you interpret the meaning of the factors). If you think there is a single factor, a histogram will do. If there are 2 factors, a 2d scatterplot will do. If there are 3 factors, make a 3d scatterplot. For 4 or more factors, make a figure with subplots, where each subplot contains 2d scatterplots (all combinations of factors).

Ambitious assignment

Do everything that the basic assignment asks you to do and – in addition:

1. The last 5 columns are (in order): Sex assigned at birth, Birth month, Annual household income, Religious identity and Profession. See whether you can build a multiple regression model using these variables to predict the total LRSP score and the score on each of the factors. How well can you predict them? What is the R^2 of your model?
2. Could you do better with a logistic model? Is there an obvious cutoff for psychopaths? If not, just use a median-split.
3. Can you substantiate the existence of clusters in this space? Are there subpopulations that are high on one factor, but low on another?