## IMC463: Machine Learning for IMC Assignment 6: Factor analysis Due: Thursday, June 2 1pm

Submit homework at the before class on Canvas. Work in your homework groups and submit one copy of the homework per group with the names of all your group members.

- 1. (27 points) What are the rewards for listening to music? A sample of American adults (age 18 or over) was asked the extent to which they agree or disagree with the following items:
  - Q28 I enjoy talking about music or bands with my friends
  - Q29 I like for other people to know which music I listen to
  - Q30 I sometimes play or share certain songs with others so that they will understand me better
  - Q31 I carefully choose the music to have on in the background when I am with my friends
  - Q32 I like to give advice and recommendations to my friends about new bands or songs
  - Q33 I like to wear t-shirts or other clothing with the logo or name of my favorite bands or musicians
  - Q34 I often "like" bands or musicians on Facebook
  - Q35 I like to have music on in the background while I'm doing other things
  - Q36 I feel energized after listening to music
  - Q37 I lose myself in the pleasure of listening to my favorite music
  - Q38 I often unwind and relax by listening to music
  - Q39 Listening to music is an escape
  - Q40 Some songs definitely affect me emotionally
  - Q41 I feel less stress after listening to music
  - Q42 When I listen to music I am worry-free
  - Q43 Part of my role among friends is to keep them informed about new music or when bands will be touring
  - Q44 Music sometimes touches me deep down
  - Q45 Some songs send shivers up my spine or give me goose bumps
  - Q46 I like to discuss my favorite bands and music on social media site
  - Q47 I enjoy following certain bands or musicians
  - Q48 I do other things better when I have music on in the background

- Q49 I sometimes feel like crying after listening to certain songs
- Q50 Being a fan of certain bands is a little like belonging to a club
- Q51 Listening to certain music leaves me with a good feeling
- Q52 My knowledge of music and bands makes me a more interesting person to others
- Q53 I can become so absorbed in music that I forget the world around me
- Q54 Listening to some songs makes me laugh out loud

Perform an exploratory factor analysis of the 27 questions to identify latent variables using either PCA or ML factor analysis. We would like to have as many factors as possible, but you should have (1) at least three items loading on each factor (three measures of each latent variable); (2) minimal large cross loadings (most less than, say .4); (3) alpha greater than 0.8. You should answer the following questions:

- (a) (3 points) How many factors should be used? Explain your rationale.
- (b) (5 points) For each factor (i.e., if you suggest 8 factors then you need to answer this part 8 times), state which items (questions) should load on the factor. List the question numbers of the questions assigned to a factor in <u>numerical order</u>. Hints: you will need to "purify" the measure by possibly dropping some of the variables that "load" on a particular factor in the EFA. Some of the considerations include discriminant validity (large cross-loadings diminish this), content validity, face validity and reliability. Don't be afraid to prune items aggressively!
- (c) (3 points) Name each factor with at most, say, five words. The factor names should accurately describe the latent variable being measured.
- (d) (3 points) Give the reliability of each scale (as measured by coefficient alpha). Does alpha improve when items are dropped?
- (e) (4 points) Provide the varimax rotated loadings of your final set of items. Also give the eigenvalues.
- (f) (3 points) Estimate scores for your factors, allowing them to be correlated (this can be done either with a promax rotation or by averaging the raw items—you might want to try both to see if you get the same conclusion!). Submit a correlation matrix. (Note: my correlations with promax range from .34 to .58, and are slightly larger with the simple averages. If yours are larger than this then you probably have items with large cross loadings, and you may encounter multicollinearity problems.)
- (g) (3 points) The time1 variable gives the length of time that the respondent spends listening to music (consumption). Regress time1 on your factors and discuss which factors are most predictive of consumption.