

MSiA 421, Data Mining

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. *Answer: Use a scree plot of PCA solution to arrive at three factors. There are 3 eigenvalues greater than 1. Substantial partial credit should be given for 2 factors. One can make a case using a scree plot for 2, but I said we want to have as many factors as possible in the instructions.*

	Eigenvalue	Difference	Proportion	Cumulative
1	14.3878122	11.8256951	0.5329	0.5329
2	2.5621172	1.4922098	0.0949	0.6278
3	1.0699073	0.3211021	0.0396	0.6674
4	0.7488053	0.0387638	0.0277	0.6951
5	0.7100415	0.0316233	0.0263	0.7214
6	0.6784182	0.1211723	0.0251	0.7466

- (b) 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 numerical order. 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! *Answer: Give credit to students if they are more aggressive in dropping items from factors as long as their alpha is greater than about .8.*

- I called it the “social-identity experience,” although anything with a social feel should receive full credit. $\alpha = .89$, questions 46, 43, 34, 33.*
- I called it the “transportive experience,” but other terms could be “timeout,” “relaxation,” etc. $\alpha = .92$, questions 38, 41, 39, 37.*
- I called it the “affect-inducing experience,” but something around an emotional or physiological response is also OK. $\alpha = .81$, questions 45, 49, 44.*

Give very substantial partial credit for a two-factor solution (without the affect-inducing experience). My correlations are $\text{corr}(\text{social}, \text{trans}) = .46$, $\text{corr}(\text{social}, \text{affect}) = .58$, $\text{corr}(\text{trans}, \text{affect}) = .34$ using a promax rotation.

- (c) Name each factor with at most, say, five words. The factor names should accurately describe the latent variable being measured.
- (d) Give the reliability of each scale (as measured by coefficient alpha). Does alpha improve when items are dropped?
- (e) Provide the varimax rotated loadings of your final set of items. Also give the eigenvalues. *Answer: Only one cross loading exceeds .4. The third eigenvalue is a bit small*

	Eigenvalue	Difference	Proportion	Cumulative
1	5.89728076	4.15491379	0.5361	0.5361
2	1.74236697	0.90728345	0.1584	0.6945
3	0.83508352	0.34406596	0.0759	0.7704
4	0.49101756	0.09671373	0.0446	0.8151
...				

	Fact1	Fact2	Fact3
V38 I often unwind/relax by listening to music	0.857	.	.
V39 Listening to music is an escape.	0.836	.	.
V41 I feel less stress after listening to music.	0.798	.	.
V37 lose myself in pleasure of listening to music.	0.739	.	.
V34 I often Like musicians on Facebook.	.	0.851	.
V46 like discuss music on social media sites	.	0.829	.
V33 wear t-shirts name of my favorite musicians.	.	0.819	.
V43 my role keep friends informed about new music	.	0.799	.
V45 songs send shivers up my spine/goose bumps.	.	.	0.795
V49 sometimes cry after listening to certain songs	.	.	0.775
V44 Music sometimes touches me deep down.	0.458	.	0.751

- (f) 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.) *Answer: Using the simple averages the correlations are a little larger: $\text{corr}(\text{social}, \text{timeout}) = .54$; $\text{corr}(\text{social}, \text{emot}) = .41$; $\text{corr}(\text{timeout}, \text{emot}) = .68$.*

	Factor1	Factor2	Factor3
Factor1	1.00000	0.46415	0.57566
Factor2	0.46415	1.00000	0.34091
Factor3	0.57566	0.34091	1.00000

- (g) 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. *Answer: The first output uses promax factor scores and the second uses the simple averages of the items. The t-statistics tell the same story: social is the more important predictor, followed by timeout. The emotional factor is not even significant.*

Variable	DF	Parameter	Standard	t Value	Pr > t
		Estimate	Error		
Intercept	1	5.83881	0.15720	37.14	<.0001
Factor1	1	2.12309	0.19128	11.10	<.0001
Factor2	1	0.93780	0.21685	4.32	<.0001
Factor3	1	0.30565	0.20070	1.52	0.1280

Variable	DF	Parameter	Standard	t Value	Pr > t	Variance
		Estimate	Error			Inflation
Intercept	1	-3.00044	0.63329	-4.74	<.0001	0
social1	1	1.84842	0.16188	11.42	<.0001	1.41276
timeout1	1	0.74799	0.22211	3.37	0.0008	2.17840
emot1	1	0.45066	0.20964	2.15	0.0318	1.85383