Thomas Kirnbauer

Multivariate Assignment 2

Overview of the Data

I conducted an exploratory factor analysis (EFA) using results from the National Survey of Student Engagement (NSSE) study on campus housing, student engagement, and persistence. The housing study was administered to all first year and senior undergraduate students at select institutions in Spring 2018. For this assignment, I examined institutional results for first year students at a large campus located in New York State. The sample size of the EFA is 491 students, which is based on a random sample of 25% of total first year respondents (1,963). There are no missing values in the dataset because the sample was selected from completed survey responses.

The goal of the housing study is for the institution to better serve their students and improve the retention of undergraduate students by understanding their experiences relating to housing. I have chosen 13 questions from the NSSE housing study, which provides valuable insight into students' living situation. The selected questions appear to be designed to measure three specific dimensions – sense of community, student wellness, and financial stress.

The sense of community question set aims to understand the students' perception relating to their feeling of safety and being valued in their community. For these items, there are four questions with a five-point Likert Scale ranging from Strongly Disagree (1) to Strongly Agree (5). An example of community-related question, is "To what extent do you agree with the following statements about where you live while attending college: I feel valued where I live." The average response for each community question is above four and there is a substantial negative skew present in the data.

The questions pertaining to wellness seek to understand how a students' living situation impacts various aspects of their health. For instance, the survey asks students "What best describes the impact your current living situation has had on the following?: Your overall emotional or mental well-being." The wellness dimension has four questions and the responses options are on a five-point Likert Scale between strong negative impact (1) to strong positive impact (5). The responses to the four questions appear normally distributed based on low skewness and kurtosis. The average response of the four variables is roughly 3.6; this could

reflect students indifference because the label for response option 3 is “no positive or negative impact.”

The final set of questions seeks to understand the level of financial stress due to cost of living and academic expenses. One question asks, “During the current school year, about how often have you done the following?: Chosen not to purchase required academic materials (books, course packs, supplies) due to their cost.” For this construct, there are five questions that use a four-point Likert Scale ranging from Never (1) to Very Often (4). Each of the financial stress items has a mean score of less than three and there is a slight positive skew. The mean for these financial questions are lower than the other questions, which makes sense as a lower score indicates that the student does not show symptoms of financial stress. Table 1 in the Appendix provides the complete descriptive statistics for 13 variables. In addition, Table 2 provides exact wording and response options for all of the survey questions.

Choosing the Number of Factors

I conducted a parallel analysis and reviewed the scree plot to determine the number of factors necessary for the factor analysis. The parallel analysis, using 100 iterations, suggests three factors is appropriate and the top three eigenvalues of factors provide a meaningful variance - 3.83, 2.53, and .71. In addition, the scree plot in Table 4 shows a considerable change in slope after the third eigenvalue. The decision to use three factors is also substantively supported given the selected items from the study. As previously alluded, the design of the survey questions should result in the 13 separate questions loading on three distinct categories – community, wellness, and financial stress.

Factor Analysis

I began the analysis by running a correlation matrix on the variables in the dataset, which can be found in Table 3. There is a consistent pattern of relatively high correlation scores among questions within items of each distinct categories (community, wellness, and financial) as shown by the highlighted values in Table 3. For instance, among the community variables, the lowest correlation is .5 between community1 and community4. The largest correlation in the dataset is .79 for wellness3 and wellness4. All of the correlations within the variables of each construct are statistically significant at $p<.001$. Conversely, the correlations between items of different dimensions are low. In fact, all but two of the correlations between the community, wellness, or

financial stress categories are below .4, which indicates the categories do not appear to be redundant measures.

I conducted a factor analysis using an unrotated, orthogonal (varimax), and oblique (promax) rotation solutions. First, I analyzed the unrotated factor analysis, and although it's results are less practical, it provides some initial understanding of the data. The unrotated solution, which is located in Table 5, demonstrates that nearly all factors load to some extent on multiple principal axis. Next, a rotation is used to better understand the output of the factor analysis.

The orthogonal (varimax) rotation results, including the communalities and uniqueness, can be found in Table 6. The communalities (h2) for the variables has a large range between .33 to .82. The items with the highest communalities are wellness4 (.82) and financial3 (.75). Whereas the variables with the highest uniqueness values (u2), which is the inverse of the communalities, is wellness2 (.67) and community1 (.53). These high uniqueness scores are not ideal to be confident with the factor analysis. The full results of the oblique (promax) rotation can be found in Table 8. The oblique rotation has the same communality scores for the individual question items.

Interpreting the Extracted Factors

For the orthogonal rotation, the three extracted factors combined to account for 62% of the cumulative variance of the observed variables. All of the financial stress items loaded solely on the second principal axis [PA], with the loadings ranging between .715 (financial5) to .864 (financial3). As a result, PA2 contained the largest proportion of the variance with 24%, and had the highest sum of square loadings with 3.18. The other two principal axis factors are not as clear because there is overlap among two of the variables - community3 and community4. The community3 variable loads .369 on PA1 as well as .696 on PA3. Whereas, the community4 has a loading of .440 on PA1 and .670 on PA3. As a result, the proportion of variance explained is lower for the associated factors, 21% for PA1 and 16% for PA3. Table 7 shows the orthogonal loadings with a cutoff value of .3; any loadings below this threshold are considered insignificant.

Table 9 shows the oblique (promax) rotation loadings with the cutoff value of .3. There are some similarities between the promax and varimax solutions. For instance, the total variance explained by the model remains as 62% of the observed values and the loadings for the third principal axis factor, PA3, are very similar. However, there are differences in the factor loadings

between the orthogonal and oblique rotation. Unlike the previous solution, the promax rotation solution resulted in each variable loading primarily on just one factor; community3 and community4 loadings no longer significantly load on two factors. Instead, for the promax solution, the community3 loading on PA1 decrease to .11 and the community4 loading decreased to .21.

There are some other notable differences among the individual variables loadings. The promax rotation resulted in several of the variable loadings increased on their primary factor. For example, community2 loading increased on PA3 from .742 (varimax) to .845 (promax). Similarly, wellness4 loading on PA1 increased from .886 (varimax) to .935 (promax). Furthermore, the factor correlations is an important consideration for the promax rotation. The factor correlation between PA1 and PA3 is .58, which is substantial. These findings provide empirical evidence that the oblique rotation is most appropriate for interpretation of the EFA.

Analyzing the Results

Comparing Observed and Reproduced Results. I compared the observed and reproduced correlations between the first variable (community1) and second variable (community2). The observed correlation value between the two variables is .56, which I found using the correlation matrix in Table 3. Next, I calculated the reproduced correlation by adding the product of the factor loadings for the two items in each of the three factors in the varimax solution. The equation to compute the reproduced correlation is as follows:

*Reproduced Correlation = (.58 * .58) + (.14 * .06) + (.34 * .49)*

The equation resulted in a reproduced correlation value of .51. To calculate the residual value of the correlation, I subtracted the observed value (.56) from the reproduced value (.51), which the result is .05. This is approximately the same as the printout, which was .05227. The residual value between the observed and reproduced correlation is slightly higher than preferred and warrants further investigation of the residuals. With that said, I do not believe the correlation residual is high enough to invalidate the analysis.

Communality. I calculated the communality for the first variable (community1) using the sum of squares of the loadings of the unrotated solution. The calculation of the sum of squares is shown in the following equation:

$$Communality = (.581)^2 + (.143)^2 + (.335)^2$$

I also calculated the communality for the same variable using the loadings from the orthogonal varimax rotation:

$$Communality = (.263)^2 + (.630)^2 + (-.060)^2$$

For both the unrotated and varimax solutions, the communality for the community1 is the same: .469. These results allow me to conclude that the rotation does not alter the extent to which the variable correlates with other items. As mentioned earlier, this particular variable had one of the lowest communalities in our dataset. But, it is still adequate to prove there is shared variance with other items.

Residual Correlation Matrix. Table 10 contains the residual correlation matrix for the oblique (promax) rotation. There are three residuals values slightly above .05 in magnitude, which are highlighted in gray on the table. These findings represent a difference between the reproduced and observed correlation matrix. Although these three residual correlation values are below a magnitude of .10, they could be problematic and warrants further investigation to better understand any potential ramifications. Nonetheless, almost all of the residuals are near 0 and there are at least four variables that primarily load on each factor in the promax solution. Thus, it is evident that the factors in the promax rotation do accurately represent the correlations of the observed variables.

Maximum Likelihood Solution

The maximum likelihood (ML) solution, using three factors, is shown in Table 11. In addition, the results of the ML with a cutoff value of .3 for the loadings is shown in Table 12. The ML solution assumes the distribution of variables and factors is normal. The maximum likelihood solution has similarities to the principal axis factor solution. The cumulative variance of the three factors remained at .61, the correlations between the factors are the same, and the sum of squares loadings for the three factors are consistent: 3.18, 2.66, and 2.16. On the other hand, there are some slights differences when looking at the individual item loadings. For six variables, the loadings on the primary factors slightly decreased – community1, community2, wellness1, wellness2, financial4, and financial5. For the other seven variables, there is slight increases in the factor loadings. It is worth noting that a major limitation of the ML solution is the assumption of normal distribution. Unfortunately, all the questions in the NSSE housing study dataset are based on a Likert-Scale and are not normally distributed.

Appendix

Table 1

Descriptive Statistics

variable	n	mean	sd	median	min	max	range	skew	kurtosis	se
community1	491	4.41	0.71	5	1	5	4	-1.36	2.95	0.03
community2	491	4.13	0.96	4	1	5	4	-1.2	1.18	0.04
community3	491	4.23	0.84	4	1	5	4	-1.23	1.76	0.04
community4	491	4.06	0.84	4	1	5	4	-0.8	0.73	0.04
wellness1	491	3.63	0.9	4	1	5	4	-0.33	-0.05	0.04
wellness2	491	3.55	1.01	4	1	5	4	-0.42	-0.33	0.05
wellness3	491	3.69	0.91	4	1	5	4	-0.42	0	0.04
wellness4	491	3.6	1.03	4	1	5	4	-0.5	-0.3	0.05
financial1	491	2.64	1.15	3	1	4	3	-0.11	-1.44	0.05
financial2	491	2.53	1.23	2	1	4	3	-0.01	-1.59	0.06
financial3	491	2.43	1.1	2	1	4	3	0.22	-1.28	0.05
financial4	491	2.26	1.17	2	1	4	3	0.34	-1.37	0.05
financial5	491	1.77	1.06	1	1	4	3	1.13	-0.11	0.05

Table 2: List of Survey Questions

Question: To what extent do you agree or disagree with the following statements about where you live while attending college? *Response options: 5=Strongly agree, 4=Agree, 3=Neither agree nor disagree, 2=Disagree, 1=Strongly disagree*	
Community1	I feel physically safe where I live.
Community2	I feel free from harassment and discrimination where I live.
Community3	I feel comfortable being myself where I live.
Community4	I feel valued where I live.

Question: What best describes the impact your current living situation has had on the following? *Response options: 5=Strong positive impact, 4=Positive impact, 3=No positive or negative impact, 2=Negative impact, 1=Strong negative impact*	
Wellness1	Your ability to succeed academically
Wellness2	Your ability to make friends in college
Wellness3	Your overall physical well-being
Wellness4	Your overall emotional or mental well-being

Question: During the current school year, about how often have you done the following? *Response options: 4=Very often, 3=Often, 2=Sometimes, 1=Never*	
Financial1	Worried about having enough money for regular expenses
Financial2	Worried about paying for college
Financial3	Chosen not to participate in an activity due to lack of money
Financial4	Chosen not to purchase required academic materials (books, coursepacks, supplies) due to their cost
Financial5	Skipped meals due to lack of funds

Table 3

Correlation Matrix (N = 491)

	community1	community2	community3	community4	wellness1	wellness2	wellness3	wellness4	financial1	financial2	financial3	financial4	financial5
community1	1												
community2	0.56	1											
community3	0.51	0.54	1										
community4	0.5	0.55	0.69	1									
wellness1	0.34	0.24	0.41	0.46	1								
wellness2	0.2	0.19	0.31	0.37	0.5	1							
wellness3	0.4	0.3	0.46	0.52	0.68	0.43	1						
wellness4	0.36	0.27	0.45	0.5	0.7	0.52	0.79	1					
financial1	-0.05	-0.12	-0.05	-0.11	-0.01	0.04	-0.04	0	1				
financial2	-0.09	-0.14	-0.08	-0.15	-0.02	0.01	-0.08	0	0.72	1			
financial3	-0.09	-0.16	-0.08	-0.1	-0.01	0.01	-0.04	-0.01	0.75	0.68	1		
financial4	-0.12	-0.13	-0.12	-0.13	0	-0.01	-0.06	-0.03	0.58	0.57	0.65	1	
financial5	-0.1	-0.14	-0.07	-0.09	-0.02	0.01	-0.04	-0.03	0.58	0.52	0.61	0.63	1

Table 4

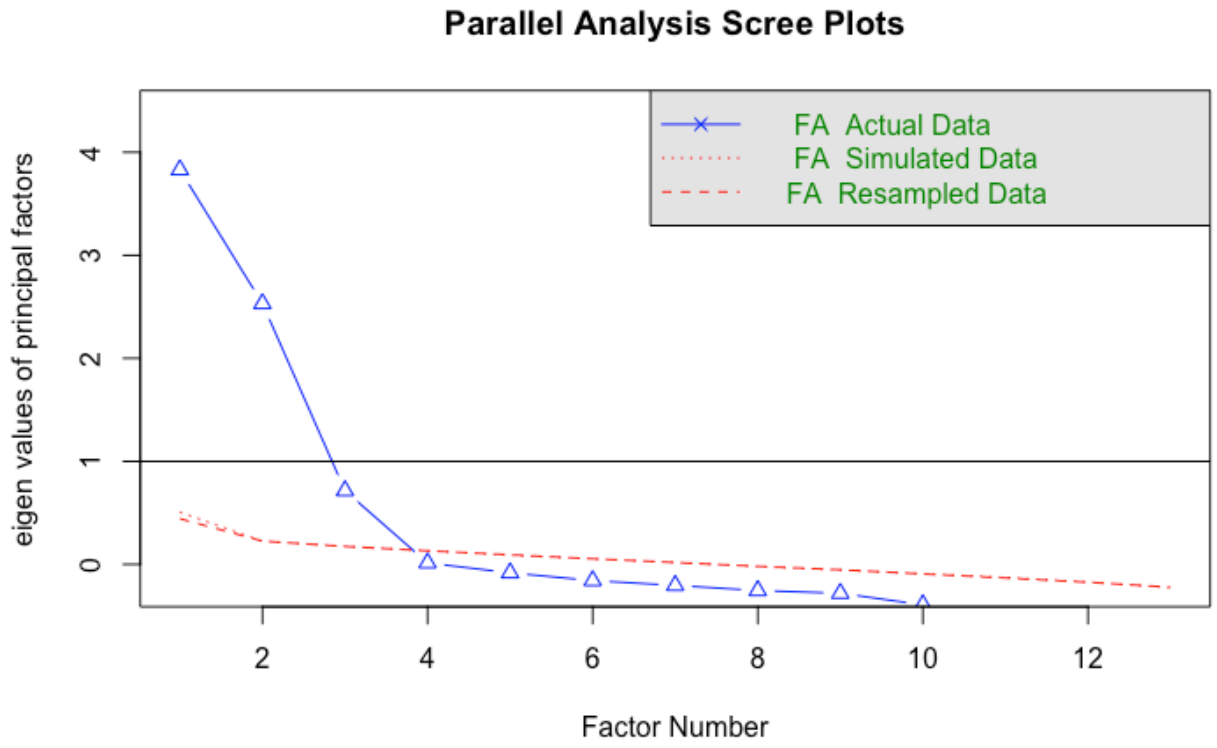
Parallel Analysis Scree Plots
FA Actual Data
FA Simulated Data
FA Resampled Data
eigen values of principal factors
0
1
2
3
4
2
4
6
8
10
12
Factor Number

Table 5
Unrotated Factory Analysis Results

Loadings (Cutoff = .3):

	PA1	PA2	PA3
community1	0.581		0.335
community2	0.582		0.491
community3	0.686		0.327
community4	0.738		
wellness1	0.657	0.343	-0.318
wellness2	0.463		
wellness3	0.728	0.321	
wellness4	0.728	0.385	-0.378
financial1	-0.358	0.766	
financial2	-0.37	0.694	
financial3	-0.38	0.775	
financial4	-0.364	0.657	
financial5	-0.332	0.634	

	PA1	PA2	PA3
SS loadings	4.049	3.054	0.916
Proportion Var	0.311	0.235	0.07
Cumulative Var	0.311	0.546	0.617

Table 6
Orthogonal Rotation (Varimax) Factor Analysis Results
Standardized loadings (pattern matrix) based upon correlation matrix

	PA2	PA1	PA3	h2	u2	com
community1	-0.06	0.26	0.63	0.47	0.53	1.4
community2	-0.12	0.14	0.74	0.58	0.42	1.1
community3	-0.04	0.37	0.7	0.62	0.38	1.5
community4	-0.09	0.44	0.67	0.65	0.35	1.8
wellness1	0.01	0.78	0.19	0.65	0.35	1.1
wellness2	0.03	0.55	0.15	0.33	0.67	1.2
wellness3	-0.04	0.81	0.25	0.72	0.28	1.2
wellness4	0.01	0.89	0.19	0.82	0.18	1.1
financial1	0.85	0	-0.02	0.72	0.28	1
financial2	0.78	-0.02	-0.07	0.62	0.38	1
financial3	0.86	0	-0.06	0.75	0.25	1
financial4	0.75	-0.01	-0.09	0.56	0.44	1
financial5	0.71	-0.01	-0.06	0.51	0.49	1

	PA2	PA1	PA3
SS loadings	3.18	2.78	2.06
Proportion Var	0.24	0.21	0.16
Cumulative Var	0.24	0.46	0.62
Proportion Explained	0.4	0.35	0.26
Cumulative Proportion	0.4	0.74	1

Table 7
Orthogonal Rotation (Varimax) Factor Analysis Results

Loadings (cutoff = .3):

	PA2	PA1	PA3
community1			0.63
community2			0.742
community3		0.369	0.696
community4		0.44	0.67
wellness1		0.784	
wellness2		0.554	
wellness3		0.81	
wellness4		0.886	
financial1	0.85		
financial2	0.785		
financial3	0.864		
financial4	0.746		
financial5	0.715		

Table 8

Oblique Rotation (Promax) Factor Analysis Results

Standardized loadings (pattern matrix) based upon correlation matrix

	PA2	PA1	PA3	h2	u2	com
community1	0.01	0.02	0.68	0.47	0.53	1
community2	-0.02	-0.17	0.85	0.58	0.42	1.1
community3	0.03	0.11	0.72	0.62	0.38	1.1
community4	-0.02	0.21	0.67	0.65	0.35	1.2
wellness1	0	0.82	-0.02	0.65	0.35	1
wellness2	0.03	0.57	0.01	0.33	0.67	1
wellness3	-0.04	0.82	0.04	0.72	0.28	1
wellness4	-0.01	0.93	-0.05	0.82	0.18	1
financial1	0.86	-0.01	0.05	0.72	0.28	1
financial2	0.79	-0.01	0	0.62	0.38	1
financial3	0.87	0.01	0.01	0.75	0.25	1
financial4	0.74	0.01	-0.04	0.56	0.44	1
financial5	0.72	0	-0.01	0.51	0.49	1

	PA2	PA1	PA3
SS loadings	3.18	2.66	2.18
Proportion Var	0.24	0.2	0.17
Cumulative Var	0.24	0.45	0.62
Proportion Explained	0.4	0.33	0.27
Cumulative Proportion	0.4	0.73	1

With factor correlations of:

	PA2	PA1	PA3
PA2	1	-0.02	-0.18
PA1	-0.02	1	0.58
PA3	-0.18	0.58	1

Table 9
Oblique Rotation (Promax) Factor Analysis Results

Loadings (cutoff = .3):

	PA2	PA1	PA3
community1			0.675
community2			0.845
community3			0.723
community4			0.666
wellness1		0.818	
wellness2		0.569	
wellness3		0.822	
wellness4		0.935	
financial1	0.858		
financial2	0.787		
financial3	0.867		
financial4	0.743		
financial5	0.716		

	PA2	PA1	PA3
SS loadings	3.18	2.66	2.18
Proportion Var	0.24	0.2	0.17
Cumulative Var	0.24	0.45	0.62
Proportion Explained	0.4	0.33	0.27
Cumulative Proportion	0.4	0.73	1

With factor correlations of:

	PA2	PA1	PA3
PA2	1	-0.02	-0.18
PA1	-0.02	1	0.58
PA3	-0.18	0.58	1

Table 10

Residual Correlation Matrix - Oblique Rotation (Promax)

	community1	community2	community3	community4	wellness1	wellness2	wellness3	wellness4	financial1	financial2	financial3	financial4	financial5
community1	0.53												
community2	0.05	0.42											
community3	-0.02	-0.03	0.38										
community4	-0.04	-0.02	0.06	0.35									
wellness1	0.02	-0.01	-0.01	-0.01	0.35								
wellness2	-0.04	0.01	0.01	0.03	0.04	0.67							
wellness3	0.02	0	-0.02	0	0	-0.05	0.28						
wellness4	0.01	0.01	0	-0.02	-0.03	0	0.03	0.18					
financial1	0.01	-0.01	0.01	-0.02	-0.01	0.02	0	0	0.28				
financial2	0.01	0.01	0	-0.02	0	0	-0.01	0.02	0.06	0.38			
financial3	0	-0.02	0	0.01	-0.01	-0.01	0.01	0	0.02	-0.01	0.25		
financial4	-0.01	0.03	-0.02	0.01	0.02	-0.01	0	-0.01	-0.05	-0.02	0	0.44	
financial5	-0.01	-0.01	0	0.02	0	0	0.01	-0.01	-0.03	-0.04	-0.01	0.09	0.49

Table 11

Maximum Likelihood Solution

Standardized loadings (pattern matrix) based upon correlation matrix

	ML2	ML1	ML3	h2	u2	com
community1	0.01	0.04	0.65	0.45	0.55	1
community2	-0.03	-0.15	0.81	0.54	0.46	1.1
community3	0.04	0.08	0.76	0.65	0.35	1
community4	-0.01	0.17	0.7	0.67	0.33	1.1
wellness1	0	0.78	0	0.62	0.38	1
wellness2	0.03	0.55	0.03	0.32	0.68	1
wellness3	-0.05	0.85	0.02	0.75	0.25	1
wellness4	-0.01	0.95	-0.06	0.83	0.17	1
financial1	0.87	-0.01	0.05	0.74	0.26	1
financial2	0.8	0	-0.01	0.64	0.36	1
financial3	0.87	0	0.01	0.75	0.25	1
financial4	0.73	0.01	-0.05	0.54	0.46	1
financial5	0.71	-0.01	0	0.5	0.5	1

	ML2	ML1	ML3
SS loadings	3.17	2.64	2.19
Proportion Var	0.24	0.2	0.17
Cumulative Var	0.24	0.45	0.62
Proportion Explained	0.4	0.33	0.27
Cumulative Proportion	0.4	0.73	1

With factor correlations of

	ML2	ML1	ML3
ML2	1	-0.02	-0.18
ML1	-0.02	1	0.59
ML3	-0.18	0.59	1

Table 12
Maximum Likelihood (ML) Solution

Loadings (cutoff = .3):

	ML2	ML1	ML3
community1			0.648
community2			0.807
community3			0.762
community4			0.704
wellness1		0.784	
wellness2		0.548	
wellness3		0.854	
wellness4		0.948	
financial1	0.868		
financial2	0.796		
financial3	0.869		
financial4	0.727		
financial5	0.705		

	ML2	ML1	ML3
SS loadings	3.175	2.603	2.157
Proportion Var	0.244	0.2	0.166
Cumulative Var	0.244	0.444	0.61