Factor Analysis of Student Mental Health Survey | Unit 4

Avinash Bunga

Master of Science in Information Systems and Business Analytics

Park University

CIS607HOF1P2024 Applied Business Forecasting

Professor: Dr. Abdelmonaem Jornaz

September 8, 2024

Factor Analysis of Student Mental Health Survey

Introduction

This analysis aims to explore the underlying factors affecting the mental health and academic experiences of IT students using data from the "Student Mental Health Survey: Online Survey on the Mental Health of IT Students" available on Kaggle. The dataset contains 21 variables capturing various aspects of students' demographics, academic engagement, mental health, and lifestyle factors. Factor analysis was used to determine common themes and patterns among these variables, providing insights into the primary factors influencing students' well-being (Ashfaq, n.d.).

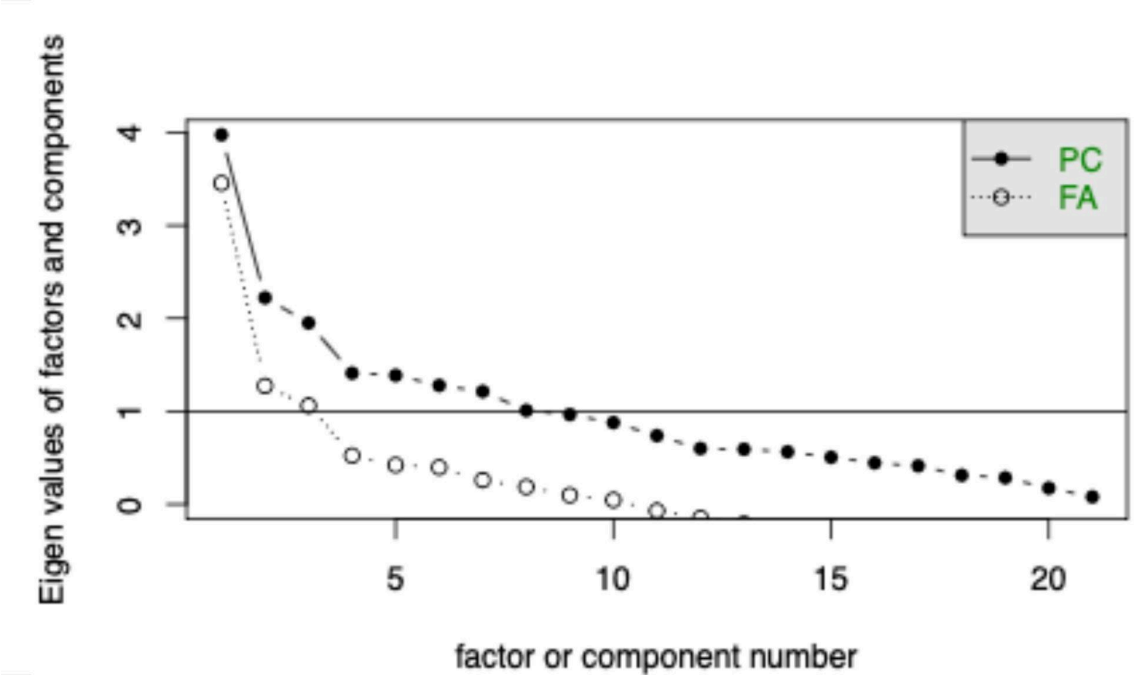
Dataset Description and Selection of Variables

The dataset includes variables such as gender, age, university, degree level, academic workload, depression, anxiety, and social relationships. Initially, all 21 variables were considered for the analysis. However, during the data preparation stage, it became evident that some variables were either redundant or did not contribute significantly to the factor structure, as indicated by their low loading values in the exploratory phase. Python was used for data cleaning, including scaling and encoding categorical variables, and R was utilized for conducting the factor analysis (Ashfaq, n.d.).

Scree Plot Analysis

A scree plot was generated to choose the number of factors to retain (see Figure 1). The scree plot displays the eigenvalues of each factor, allowing for the identification of the "elbow point," where the eigenvalues begin to level off. Based on the scree plot results, three factors were retained for further analysis as they accounted for the most significant variance and showed distinct separation from subsequent factors (Sanderson, 2023).

Figure 1: Scree Plot



Explanation of Figure 1: The scree plot of the factor analysis helps determine the optimal number of factors. The plot indicates an “elbow” at the third factor, suggesting that the first three factors capture the most meaningful variance in the data (Sanderson, 2023).

Factor Analysis Results

After determining the number of factors, the analysis proceeded with extracting and rotating three factors using Varimax rotation to maximize interpretability. The final factor loadings for significant variables are presented in Table 1 (Murphy, 2021).

Table 1: Factor Loadings of Significant Variables

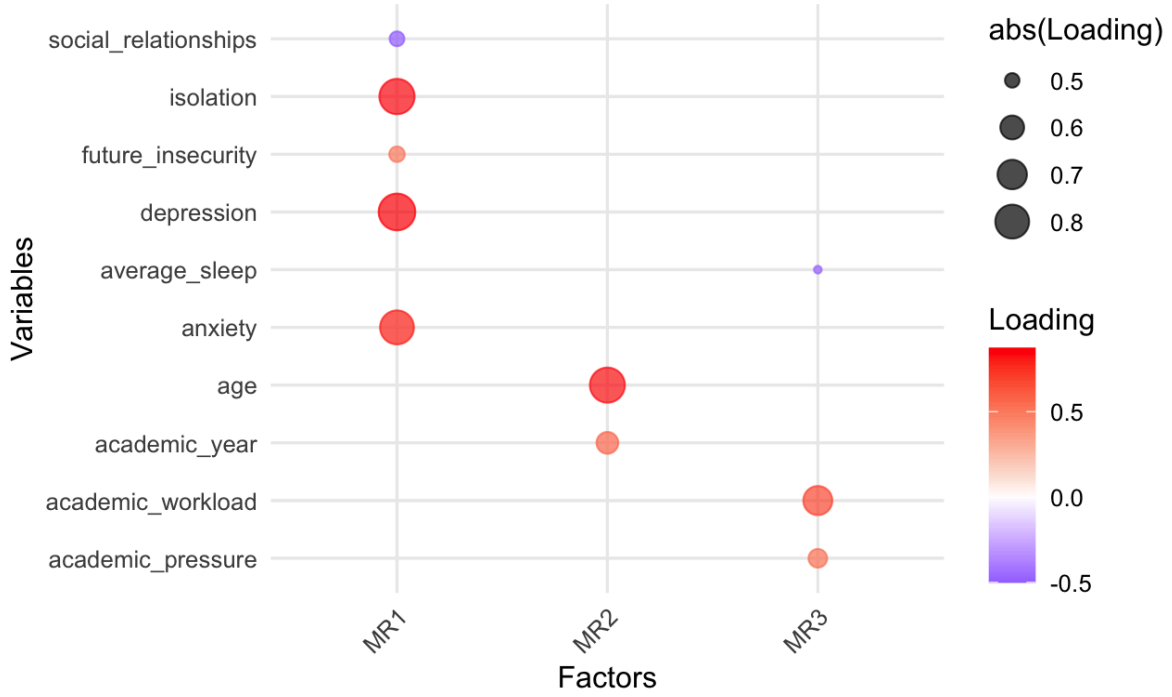
Factor	Variables with High Loadings	Interpretation
Factor 1: Mental Health and Emotional Well-being (MR1)	Depression (0.873)	This factor captures variables strongly linked to students' mental health, highlighting emotional challenges such as anxiety, depression, and feelings of isolation.
	Anxiety (0.801)	
	Isolation (0.835)	
	Future Insecurity (0.507)	
Factor 2: Academic Engagement and Age (MR2)	Age (0.829)	Represents academic engagement, reflecting students' progression through their studies and maturity. Older students and those in advanced years show distinct academic engagement patterns.
	Academic Year (0.570)	
Factor 3: Academic Workload and Lifestyle (MR3)	Academic Workload (0.691)	This factor combines the impact of academic demands on students' lifestyles, particularly highlighting how increased workload and pressure can negatively affect sleep patterns.
	Academic Pressure (0.534)	
	Average Sleep (-0.478)	

(Murphy, 2021).

Visualization of Factor Loadings

To better understand the contributions of each variable to the identified factors, a dot plot of the significant factor loadings was generated (see Figure 2). This plot highlights the variables with loadings above a threshold of 0.4, focusing on the strongest associations between variables and factors (R Coder, n.d.).

Figure 2: Significant Factor Loadings Plot



Explanation of Figure 2: illustrates the significant factor loadings, with dot sizes and colors representing the strength and direction of the loadings. The plot highlights the key variables that define each factor, making it easier to interpret the groupings (R Coder, n.d.).

Practical Application of Factor Scores

The identified factors can be used to forecast various aspects of student well-being. For example, high scores on Factor 1 (Mental Health and Emotional Well-being) may indicate a higher risk of emotional distress, suggesting the need for targeted mental health support. Similarly, students with high scores on Factor 3 (Academic Workload and Lifestyle)

could be experiencing high academic stress, pointing towards interventions like workload management and counseling (StatisticsSolutions, n.d.).

Model Quality Discussion

- **KMO Test**: The overall KMO value was 0.59, which indicates marginal adequacy for factor analysis. While slightly below the ideal threshold, the results still provide meaningful insights into the factors affecting students.
- **Bartlett's Test**: The test was highly significant ($\chi^2 = 561.14$, $p < 0.001$), confirming that the correlations among the variables are sufficient to perform factor analysis.
- **Variance Explained**: The three factors together explain 30.7% of the total variance, with the first factor contributing the most, emphasizing the importance of mental health among students (Rpubs, n.d.).

Conclusion

The factor analysis successfully identified three key dimensions of student experiences: mental health, academic engagement, and academic workload. These factors provide valuable insights into the challenges faced by IT students and highlight areas where targeted interventions could improve student well-being and academic success.

References

Ashfaq, A. (n.d.). *Student Mental Health Survey*. Kaggle. Retrieved September 8, 2024, from https://www.kaggle.com/datasets/abdullahashfaqvirk/student-mental-health-survey

Murphy, P. (2021, April 21). *Exploratory Factor Analysis*. Rpubs. https://rpubs.com/pjmurphy/758265

R Coder (n.d.). *Dot plot in R*. Retrieved September 8, 2024, from https://r-coder.com/dot-plot-r/

Rpubs (n.d.). *Factor Analysis*. Retrieved September 8, 2024, from https://rpubs.com/jaelison/135029

StatisticsSolutions (n.d.). *Comprehensive Guide to Factor Analysis*. Retrieved September 8, 2024, from https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/factor-analysis/

Sanderson, S. P. (2023, October 24). *Creating a Scree Plot in Base R*. Spsanderson. https://www.spsanderson.com/steveondata/posts/2023-10-24/index.html