**UDPS2233 MULTIVARIATE ANALSYSIS**

**ASSIGNMENT 2**

1.0 Cleaning data and Determine Number of Factor

The SMAData.xlsx contains answers for 22 questions in a questionnaire that tests the smart phone addiction and it contains 289 observations. After analysing the data, some missing values has been found in observation ID 81 and ID 154, therefore the row of ID 81 and ID 154 were deleted, so the dataset is remaining only 287 observations.

After cleaning data, scree-plot of parallel analysis method was used to determine the number of factors in Rstudio by using fa.parallel( ) function; parallel analysis suggested that 4 factors and components should be used.

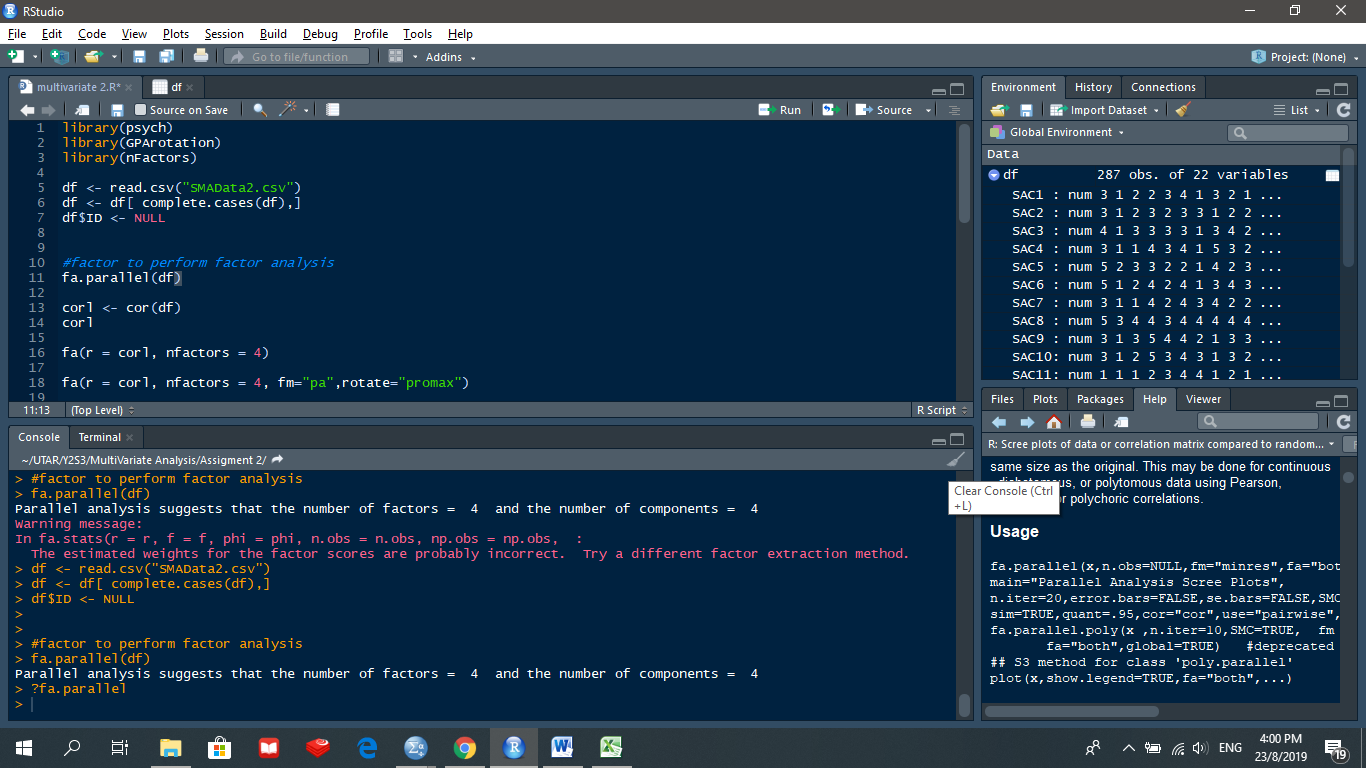


Figure 1. The result of parallel analysis

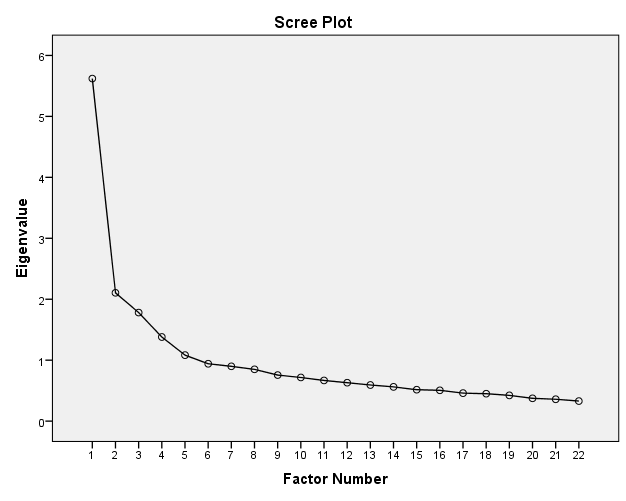


Figure 2. Scree Plot

In Figure 2, the scree plot with an arrow indicating the point of inflexion on the curve. Thus, 4 number of factors are considered to run in the SPSS with two different methods.

2.0 Factor Analysis

2.1 Covariance structure

2.2 Kaiser-Meyer-Olkin (KMO) and Bartlett’s Test (measures the strength of relationship among the variables)

|  |  |  |
| --- | --- | --- |
| **KMO and Bartlett's Test** | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .859 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 1767.275 |
| df | 231 |
| Sig. | .000 |

Table 1. Result of KMO and Bartlett’s Test

In Table 1, the result shows that the value of KMO is 0.859 which is acceptable, it indicates that the responses given with the sample are adequate.

For Bartlett’s test, the null hypothesis states that the covariance matrix is an identity matrix. The result shows that the significant value is lesser than 0.001, so null hypothesis is rejected because there is sufficient evidence to conclude that the covariance matrix is an identity matrix and it’s indicate the data is suitable to do factor analysis.

2.3 Principal Component Factor Analysis, including Varimax rotation

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Total Variance Explained** | | | | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 5.621 | 25.551 | 25.551 | 5.621 | 25.551 | 25.551 | 3.137 | 14.259 | 14.259 |
| 2 | 2.105 | 9.569 | 35.120 | 2.105 | 9.569 | 35.120 | 2.717 | 12.352 | 26.611 |
| 3 | 1.781 | 8.095 | 43.215 | 1.781 | 8.095 | 43.215 | 2.583 | 11.739 | 38.350 |
| 4 | 1.380 | 6.275 | 49.490 | 1.380 | 6.275 | 49.490 | 2.451 | 11.140 | 49.490 |
| 5 | 1.083 | 4.923 | 54.413 |  |  |  |  |  |  |
| 6 | .941 | 4.279 | 58.692 |  |  |  |  |  |  |
| 7 | .899 | 4.086 | 62.778 |  |  |  |  |  |  |
| 8 | .849 | 3.861 | 66.639 |  |  |  |  |  |  |
| 9 | .755 | 3.434 | 70.073 |  |  |  |  |  |  |
| 10 | .716 | 3.255 | 73.328 |  |  |  |  |  |  |
| 11 | .668 | 3.036 | 76.365 |  |  |  |  |  |  |
| 12 | .630 | 2.865 | 79.229 |  |  |  |  |  |  |
| 13 | .592 | 2.693 | 81.922 |  |  |  |  |  |  |
| 14 | .562 | 2.553 | 84.475 |  |  |  |  |  |  |
| 15 | .515 | 2.342 | 86.817 |  |  |  |  |  |  |
| 16 | .506 | 2.299 | 89.116 |  |  |  |  |  |  |
| 17 | .459 | 2.087 | 91.203 |  |  |  |  |  |  |
| 18 | .451 | 2.049 | 93.252 |  |  |  |  |  |  |
| 19 | .423 | 1.921 | 95.173 |  |  |  |  |  |  |
| 20 | .374 | 1.701 | 96.874 |  |  |  |  |  |  |
| 21 | .359 | 1.632 | 98.506 |  |  |  |  |  |  |
| 22 | .329 | 1.494 | 100.000 |  |  |  |  |  |  |
| Extraction Method: Principal Component Analysis. | | | | | | | | | |

Table 2. Result of Total Variance Explained

Before extraction, SPSS has identified 22 linear components within the dataset. Due to the result we obtained from the parallel analysis, we set the number of factors to 4 instead of allowing the SPSS helps us to identify. After principal component analysis is applied, it is observed that the result before extraction is same with the result after extraction except for the values for discarded factors are ignored. According to Table 2, rotation has the effect of optimizing the factor structure and one consequence for these data is that the relative importance of the four factors. Before rotation, component 1 accounted for considerably more variance than the remaining three components. For example, component 1 explains 25.551% of total variance while component 4 explains 6.275% of total variance. However, after rotation, component 1 accounts for only 14.259% of variance (compared to 12.352%, 11.739%, and 11.140% respectively). In overall, total of 49.49% of variability can be explained by using the 4 components.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rotated Component Matrixa** | | | | |
|  | Component | | | |
| 1 | 2 | 3 | 4 |
| SAC16 | .760 | -.054 | .194 | .165 |
| SAC20 | .667 | .082 | .050 | .100 |
| SAC19 | .662 | .281 | .071 | -.037 |
| SAC14 | .623 | .300 | .101 | .212 |
| SAC13 | .490 | .001 | .441 | .008 |
| SAC7 | .391 | .177 | .107 | .243 |
| SAC8 | -.333 | .676 | .072 | .107 |
| SAC21 | .345 | .672 | .010 | .184 |
| SAC22 | .108 | .629 | .186 | -.051 |
| SAC15 | .227 | .575 | .098 | .011 |
| SAC12 | .046 | .572 | -.142 | .456 |
| SAC18 | .388 | .497 | .091 | .104 |
| SAC3 | .008 | .077 | .794 | .061 |
| SAC9 | .030 | .162 | .715 | .057 |
| SAC17 | .405 | -.011 | .627 | .154 |
| SAC4 | .154 | .178 | .586 | .421 |
| SAC1 | .136 | .006 | .520 | .453 |
| SAC2 | -.008 | .117 | .096 | .744 |
| SAC5 | .075 | -.008 | .191 | .577 |
| SAC10 | .210 | .465 | .033 | .559 |
| SAC11 | .421 | .050 | -.121 | .491 |
| SAC6 | .125 | .048 | .181 | .432 |
| Extraction Method: Principal Component Analysis.   Rotation Method: Varimax with Kaiser Normalization.  Table 3. Rotated Component Matrix | | | | |

Rotated Component Matrix shows you the factor loadings for each variable and the high-lighted values are the strongest loading between its variable. Based on the factor loadings,

Component 1 – SAC 7, SAC 13, SAC 14, SAC 16, SAC 19, SAC 20

Component 2 - SAC 8, SAC 12, SAC 15, SAC 18, SAC 21, SAC 22

Component 3 - SAC 1, SAC 3, SAC 4 SAC 9, SAC 17

Component 4 – SAC 2, SAC 5, SAC 6, SAC 10, SAC 11

2.4 Maximum Likelihood Factor Analysis, including Varimax rotation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Total Variance Explained** | | | | | | |
| Factor | Initial Eigenvalues | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 5.621 | 25.551 | 25.551 | 2.586 | 11.755 | 11.755 |
| 2 | 2.105 | 9.569 | 35.120 | 2.112 | 9.599 | 21.353 |
| 3 | 1.781 | 8.095 | 43.215 | 2.090 | 9.502 | 30.855 |
| 4 | 1.380 | 6.275 | 49.490 | 1.766 | 8.028 | 38.883 |

Table 4. Total Variance Explained

According to the table above by using maximum likelihood estimate with varimax rotation total of 38.89% of variability can be explained by using 4 components. After varimax rotation is applied, the total variance explained by the 4 components has been decreased from 49.49% to 38.883%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rotated Factor Matrixa** | | | | |
|  | Factor | | | |
| 1 | 2 | 3 | 4 |
| SAC16 | .744 | -.058 | .189 | .170 |
| SAC14 | .586 | .290 | .116 | .196 |
| SAC19 | .570 | .253 | .084 | .023 |
| SAC20 | .531 | .101 | .086 | .128 |
| SAC13 | .421 | .024 | .376 | .027 |
| SAC7 | .345 | .190 | .123 | .193 |
| SAC21 | .337 | .654 | .033 | .158 |
| SAC12 | .060 | .534 | -.070 | .391 |
| SAC8 | -.246 | .534 | .064 | .126 |
| SAC22 | .133 | .508 | .148 | .013 |
| SAC15 | .221 | .462 | .089 | .065 |
| SAC18 | .343 | .419 | .105 | .143 |
| SAC3 | .030 | .066 | .716 | .062 |
| SAC9 | .061 | .138 | .603 | .071 |
| SAC17 | .399 | -.005 | .573 | .151 |
| SAC4 | .181 | .185 | .547 | .365 |
| SAC1 | .147 | .024 | .473 | .403 |
| SAC2 | .015 | .120 | .128 | .656 |
| SAC10 | .225 | .418 | .067 | .526 |
| SAC11 | .351 | .087 | -.031 | .403 |
| SAC5 | .113 | .072 | .199 | .383 |
| SAC6 | .139 | .095 | .175 | .285 |
| Extraction Method: Maximum Likelihood.  Rotation Method: Varimax with Kaiser Normalization.  Table 5. Rotated Factor Matrix | | | | |

Rotated Factor Matrix shows you the factor loadings for each variable and the high-lighted values are the strongest loading between its variable. Based on the factor loadings,

Component 1 – SAC 7, SAC 13, SAC 14, SAC 16, SAC 19, SAC 20

Component 2 - SAC 8, SAC 12, SAC 15, SAC 18, SAC 21, SAC 22

Component 3 - SAC 1, SAC 3, SAC 4 SAC 9, SAC 17

Component 4 – SAC 2, SAC 5, SAC 6, SAC 10, SAC 11

2.4 Comparison between Principal Component Factor Analysis (Varimax rotation) and Maximum Likelihood Factor Analysis (Varimax rotation)

|  |  |  |
| --- | --- | --- |
|  | Principal Component Analysis | Maximum Likelihood |
| Factor 1 | SAC 7 | SAC 7 |
| SAC 13 | SAC 13 |
| SAC 14 | SAC 14 |
| SAC 16 | SAC 16 |
| SAC 19 | SAC 19 |
| SAC 20 | SAC 20 |
| Factor 2 | SAC 8 | SAC 8 |
| SAC 12 | SAC 12 |
| SAC 15 | SAC 15 |
| SAC 18 | SAC 18 |
| SAC 21 | SAC 21 |
| SAC 22 | SAC 22 |
| Factor 3 | SAC 1 | SAC 1 |
| SAC 3 | SAC 3 |
| SAC 4 | SAC 4 |
| SAC 9 | SAC 9 |
| SAC 17 | SAC 17 |
| Factor 4 | SAC 2 | SAC 2 |
| SAC 5 | SAC 5 |
| SAC 6 | SAC 6 |
| SAC 10 | SAC 10 |
| SAC 11 | SAC 11 |
| Total variance explained | 49.9% | 38.89% |

Table 6. Comparison between PCA and MLE

Table 6 shows that using Principle Component Analysis and Maximum Likelihood the same variables will fall in the same components respectively, which means that the variables are in the same group. Since the variance explained by Principal Component Analysis (49.49%) is higher that Maximum Likelihood (38.89%), so factor analysis using Principal Component Analysis with Varimax Rotation is much more suitable.