Problem 1: Journal Article Review

Review and critique the journal article for the following questions related to factor analysis below.

See article on Psychometric Properties of Attitude towards e-Learning Scale among Nursing Students in Moodle

- How are they applying Factoring Analysis?
- What kind of factor rotation do they use?
- How many factors do they concentrate on in their analysis? How did they arrive at these number of factors?
- Explain the breakdown of the factors and the significance of their names.
- How do they evaluate the stability of the components (i.e. factorability)?
- Do they use these factors in later analysis, such as regression? If so, what do they discover?
- What overall conclusions does Factor Analysis allow them to draw?

1. How are they applying factoring analysis?

Answer:

They are trying to construct a scale that shows the attitude of Filipino nursing students who have transitioned to e-learning due to COVID-19. This is done so they can assess the attitude of those nursing students to achieve a smooth transition towards e-learning. They use exploratory factor analysis using PCA here to assess the scale for construct validity.

2. What kind of factor rotation do they use?

Answer:

Exploratory factor analysis using principal component analysis with varimax rotation to figure out the scales construct validity.

3. How many factors do they concentrate on in their analysis? How did they arrive at these number of factors?

Answer:

The e-learning scale consisted of 11 sub-scales. Using PCA and exploratory factor analysis with varimax rotation, the construct validity was determined. Components with eigen values greater than 1.0 was used to examine the factors in each of the 11 sub-scales. For the factor loadings, a threshold on 0.6 and more was kept to retain items in the scale. Evaluation of PCA led to the removal of 2 scales, leaving them with 9. The remaining items on the scale were loaded on one factor. The factor loadings ranged from –0.907 to 0.893.

4. Explain the breakdown of the factors and the significance of their names.

Answer:

The factors considered are used to study whether e-learning will be applicable for nursing students.

Factors:

- **I am interested in studying courses that utilize e-learning**: this would be a helpful factor to consider because the interest of students towards e-learning might contribute towards the scale.
- **I think that e-learning promotes my learning experiences: apart from one's interest in e-learning itself, their opinion towards whether it would help them is significant.**
- **Presenting courses on the internet makes learning more efficient**: presenting courses online could make it similar to courses in-person, hence it is an important factor.
- **I intend to use e-learning tools during the semester if available**: a persons willingness to use e-learning during a semester is a factor that is important because if they're not comfortable using it, it might not be useful.
- **I am positive about e-learning**: a person's attitude towards e-learning is significant because if they associate negatively with it, it could have an impact.
- **E-learning environment needs advanced technical knowledge on computer use**: a person's perception of how e-learning works and whether they think it needs knowledge about computers might or might not be an important factor. It depends on a person's comfortableness with using computers. It is NOT selected in factor analysis.
- **I would prefer to have courses on the internet rather than in the classroom or face-to-face**: a person's preference on whether they want to attend classes online is significant to determine the overall goal.
- **Online learning is more comfortable and enjoying to me**: this would correlate with the previous factor, which is significant. If they prefer it and whether or not they enjoy it are two different things that are important to consider.
- **E-learning is a favorable alternative to the pen-paper based system**: some people might prefer online exams than pen and paper since it could be faster. This would correlate with the other factors in the scale and hence is significant.
- **E-learning is not an efficient learning method**: whether or not a person thinks e-learning is efficient could affect their attitude towards it.
- **Over-all, I prefer e-learning and I believe that it is better than traditional method of learning**: this factor is said to capture overall feeling towards e-learning, but can fail to capture certain things which could be missed in their overall consideration. This is NOT selected during factor analysis.

5. How do you evaluate the stability/factorability of the components?

Answer:

Components with eigenvalues greater than 1.0 were used to examine the factors in each subscale. Kaiser-Meyer-Olkin (KMO) test of 0.6 was used to assess the adequacy of samples. Bartletts test of sphericity was used to assume factorability of correlation matrix with significance set to 0.05.

KMO test gave a value of 0.900. Barrelettes' test gave a p-value of 0.000. To assess the factorability of the matrix, it was recommended that the adequacy of samples in the factor analysis has a value of 0.60. This criterion was met for the results of factor analysis of the instrument.

6. Do they use these factors in later analysis, such as regression? If so, what do they discover?

Answer:

No, they do not use regression or any machine learning techniques. They have devised an instrument which is capable of determining the attitude of nursing students towards e-learning. The paper discusses the validity of the instrument and how they proved it to be valid, what tests did they run etc. The instrument itself is predictive. Factor analysis was used to

Problem 2 Principal Component Analysis

The data given in the file 'Big5.csv' are 5-point Likert items taken from the Big Five Personality Test web-based personality assessment. Techniques, such as Principal Component Analysis (PCA), can be used to determine different types of personalities. There are 19,719 subjects in the file and 50 variable items.

A) How many components are determined from the scree plot using the knee method?

Answer:

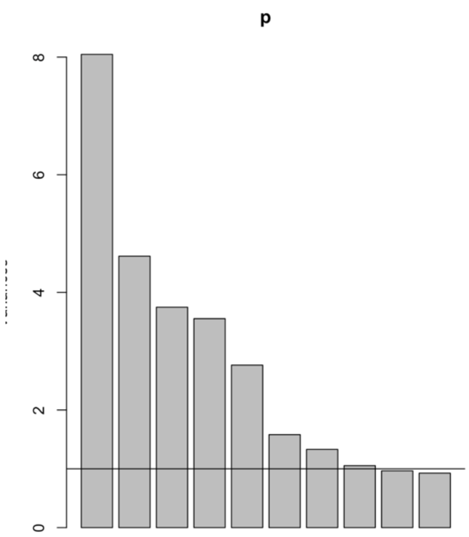
```
#center and scale the data
p = prcomp(big5, center=T, scale=T)
p

#Check Scree Plot
plot(p)
abline(1, 0)

#components with eigen values > 1
#knee is closer to 8

#Check PCA Summary Information
summary(p)
print(p)

biplot(p)
```



Using the prcomp function, we scale and skew the data to generate the knee plot. We keep the variance threshold as 1. We can see that close to 8 components p-values have variance > 1. But using the knee method, we can see that there is a drop in values from the 5^{th} and. 6^{th} bar giving us a knee at ~6. Hence, using the knee method we can see that the number of components we must select is approx. 6.

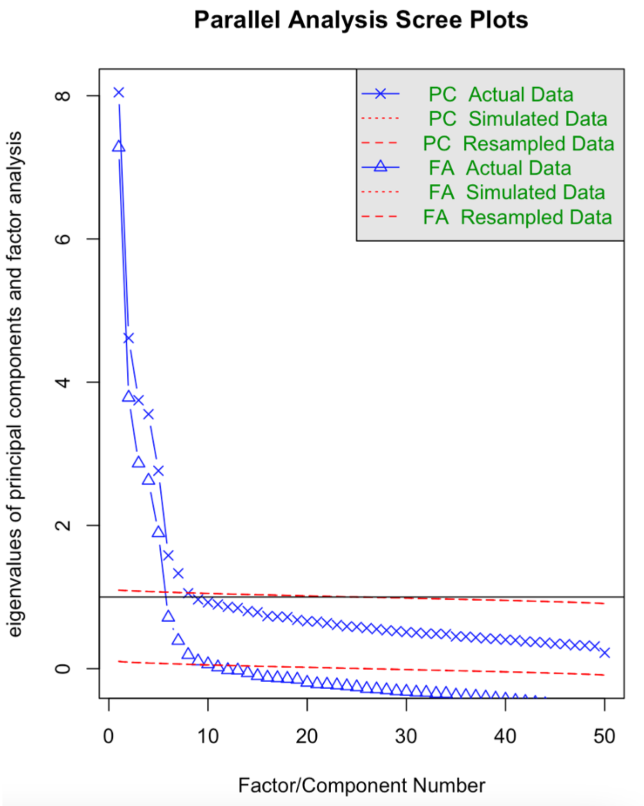
B) What number of components would you using the eigenvalue method?
Answer:

Using parallel analysis with Monte Carlo Simulations, we get the following output:

```
library(psych)

comp <- fa.parallel(big5)
comp
```

Outputs:



```
> comp <- fa.parallel(big5)
Parallel analysis suggests that the number of factors =  10  and the number of components =  7
> comp
Call: fa.parallel(x = big5)
Parallel analysis suggests that the number of factors =  10  and the number of components =  7

 Eigen Values of 
   Original factors Resampled data Simulated data Original components Resampled components Simulated components
1              7.28           0.10           0.10                8.05                 1.09                 1.10
2              3.79           0.09           0.09                4.62                 1.09                 1.09
3              2.87           0.08           0.08                3.75                 1.08                 1.08
4              2.63           0.08           0.08                3.55                 1.08                 1.08
5              1.89           0.07           0.07                2.76                 1.07                 1.07
6              0.72           0.07           0.07                1.58                 1.07                 1.07
7              0.39           0.06           0.07                1.33                 1.06                 1.06
8              0.19           0.06           0.06                1.05                 1.06                 1.06
9              0.10           0.06           0.06                0.97                 1.05                 1.05
10             0.06           0.05           0.05                0.93                 1.05                 1.05
```

From the outputs above, we see that the recommended number of components is 7 and the number of factors to consider is 10, according to the Monte Carlo simulations. But, according to the eigen values (of the original components) we see that only 8 components have values >1. Hence, using the eigen value method, we select number of components to be 8.

C) Based upon your answers from parts A and B, what number of components would you wish to start with for the model? Run the PCA model.
Answer:

Based upon the two types of tests done: one with the Knee method and one with using the eigen values, I'm picking the **number of components to be 6**, in order to run the PCA model with varimax rotation.

Code to run PCA model:

```
p2 = psych::principal(big5, rotate="varimax", nfactors=6, scores=TRUE)
p2

p3<-print(p2$loadings, cutoff=.47, sort=T)


#PCAs Other Available Information

ls(p2)

p2$values
p2$communality
p2$rot.mat
```

D) For the number of components in part C, give the formula for the first component.
Answer:

Formula for RC1 First Component

(0.692)*E1 + (-0.734)*E2 + (0.657)*E3 + (-0.754)*E4 + (0.742)*E5 + (-0.632)*E6 + (0.745)*E7 + (-0.627)*E8 + (0.653)*E9 + (-0.698)*E10

E) Give a brief interpretation of the components after rotation. What do these components mean? What names might you give for each of the components?

Answer:

Component 1 captures information from E1-E10, which talks about things related to a person's social life. I would call RC1 as "Social Skills".

Component 2 captures information from N1-N10, excluding N4. The N variable talks about a person's mood. I would name RC2 as "Mood".

Component 5 captures information from A2-A9, excluding A1, A3 and A10. The A variables talks about a person's feelings. I would name RC5 as "Feelings".

Component 3 captures information from C1-C10, excluding C3. The C variables talks about a person's actions. I would name RC3 as "Actions".

Component 4 captures information from O1-O10, excluding O9. The O variables talk about a person's thoughts and personal opinions about themselves. I would name RC4 as personal opinions.

F) What are the highest and lowest scores for each principal component conducted in Part C?

Answer:

```
#Calculating scores

scores <- p2$scores

cor(scores)

summary(scores)

scores_1 <- scores[,1]

min_score <- min(scores_1)
min_score

max_score <- max(scores_1)
max_score

scores_2 <- scores[,2]
min_score <- min(scores_2)
min_score

max_score <- max(scores_2)
max_score

scores_3 <- scores[,3]
min_score <- min(scores_3)
min_score

max_score <- max(scores_3)
max_score

scores_4 <- scores[,4]
min_score <- min(scores_4)
min_score

max_score <- max(scores_4)
max_score

scores_5 <- scores[,5]
min_score <- min(scores_5)
min_score

max_score <- max(scores_5)
max_score

scores_6 <- scores[,6]
min_score <- min(scores_6)
min_score

max_score <- max(scores_6)
max_score
```

Values:

COMPONENT	MIN SCORE	MAX SCORE
RC1	-2.858513	2.895168
RC2	-3.921192	2.708987
RC5	-4.471487	2.324091
RC3	-3.434239	2.893127
RC4	-4.731591	2.449809
RC6	-9.543987	6.280175

G) Finally, run a common factor analysis on the same data. Is there a difference between the Principal Component Analysis and the factor analysis? Does the factor analysis change your ability to interpret the results practically?

<u>Answer</u>:

<u>Code</u>:

```
#Conducting Factor Analysis

fit = factanal(big5, 5)
print(fit$loadings, cutoff=.4, sort=T)
```

<u>Output</u>:

```
> fit = factanal(big5, 5)
> print(fit$loadings, cutoff=.4, sort=T)

Loadings:
    Factor1 Factor2 Factor3 Factor4 Factor5
E1   0.669
E2  -0.677
E3   0.655
E4  -0.697
E5   0.733
E6  -0.567
E7   0.739
E8  -0.552
E9   0.606
E10 -0.650
N1           0.670
N3           0.597
N5           0.539
N6           0.741
N7           0.723
N8           0.751
N9           0.708
N10          0.611
A2                   0.529
A4                   0.790
A5                  -0.651
A6                   0.595
A7                  -0.610
A8                   0.578
A9                   0.708
C1                           0.599
C2                          -0.528
C4                          -0.556
C5                           0.624
C6                          -0.582
C7                           0.535
C9                           0.624
O1                                   0.576
O2                                  -0.537
O3                                   0.533
O5                                   0.607
O6                                  -0.504
O8                                   0.539
O10                                  0.679
N2          -0.489
N4
A1                  -0.425
A3                  -0.401
A10
C3                           0.401
C8                          -0.476
C10                          0.469
O4                                  -0.464
O7                                   0.490
O9

               Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings      4.993   4.600   3.756   3.272   3.170
Proportion Var   0.100   0.092   0.075   0.065   0.063
Cumulative Var   0.100   0.192   0.267   0.332   0.396
>
```

For PCA, I’ve used 6 components and for factor analysis I’ve used 5 factors (in Monte Carlo simulation 5 factors have eigenvalues greater than 1, hence why I chose 5). Here, all factors correlate with some or all variables in each field (E, A etc). In PCA, component 6 does not correlate with any of the variables.

Factor 1 captures all the E variables, similar to component 1. Factor 2 captures all the N variables, but PCA excludes N4 (I seldom feel blue). Factor 3 captures all A variables excluding A10, but Component 5 excludes A1 and A3. Factor 4 captures all the C variables, but PCA excludes C3. Factor 5 captures all the O variables, but excludes O9, similar to PCA. Factor analysis captures the details more than PCA, so it increases ability to interpret the results.