STAT6108 Assignment 3 Suggested Solution 2nd Term, 2019-2020

(a)

(1) Levels of Measurement: Input data for an exploratory factor analysis should be continuous, i.e. in interval or ratio scale at best, or in ordinal scale if the number of categories is larger than or equal to 5. The current items are coded in a 4-point Likert scale, which is marginally adequate for EFA.

(2) Sample Size: The current scale has 10 items and the sample size is 89, N/p ratio=89/10=8.9, which is acceptable as N/p>5.

(3) Data Dependency: If all variables are independent, the data are inappropriate for EFA. To check this, we use (i) Bartlett's test of sphericity and (ii) KMO index of sampling adequacy:

(i) Bartlett's test of sphericity:

```
> # Bartlett's test and KMO measure of sampling adequacy
> cortest.bartlett(mycor, n=nobs)
$chisq
[1] 262.3824

$p.value
[1] 1.816464e-32

$df
[1] 45
```

Barlett's test of sphericity is significant, X^2(45)=262.38, p=.00<.05, indicating that not all variables are independent. Hence, the variables are dependent enough for EFA.

(ii) KMO index of sampling adequacy:

```
> KMO(mycor)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = mycor)
Overall MSA =  0.72
MSA for each item =
  V1   V2   V3   V4   V5   V6   V7   V8   V9  V10
0.74 0.76 0.77 0.68 0.68 0.71 0.78 0.69 0.75 0.64
```

KMO=0.72 > 0.7, indicating that the variables share adequate amount of common variance.

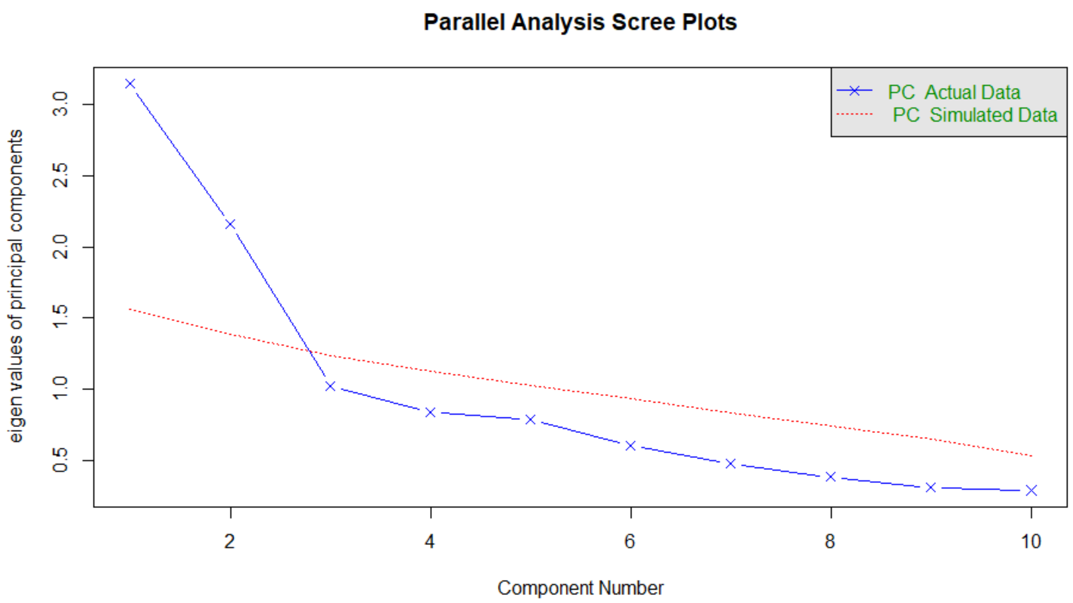
Considering all the factors above, the current data set is reasonably appropriate for EFA.

(b)

```
> # Complete factor solutions using principal component extraction
> principal(mycor, nfactors=p, n.obs=nobs, rotate="none")
Principal Components Analysis
Call: principal(r = mycor, nfactors = p, rotate = "none", n.obs = nobs)
Standardized loadings (pattern matrix) based upon correlation matrix
     PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9  PC10 h2       u2 com
V1  0.71 -0.39  0.18 -0.24 -0.25 -0.04  0.14  0.11 -0.40 -0.06  1  4.4e-16 3.2
V2  0.64  0.46  0.08  0.11 -0.24 -0.26 -0.42  0.05 -0.01  0.23  1  2.2e-15 3.9
V3  0.56 -0.33  0.12  0.04  0.68  0.02 -0.21  0.23  0.00 -0.08  1  1.1e-16 3.1
V4  0.54  0.64  0.26  0.00  0.09 -0.09 -0.01 -0.33  0.00 -0.31  1  1.6e-15 3.5
V5  0.52  0.23 -0.63  0.01 -0.18  0.44 -0.15  0.08  0.00 -0.15  1  8.9e-16 3.6
V6  0.61 -0.49  0.30  0.13 -0.35 -0.01  0.09  0.15  0.34 -0.12  1 -2.0e-15 4.3
V7  0.57 -0.06 -0.60  0.06  0.13 -0.46  0.27 -0.02  0.06  0.03  1  5.6e-16 3.5
V8  0.22  0.68  0.18  0.53  0.07  0.16  0.31  0.20 -0.08  0.09  1  8.9e-16 3.3
V9  0.70 -0.45  0.05  0.11  0.12  0.29  0.06 -0.36  0.02  0.25  1 -2.2e-16 3.3
V10 0.34  0.56  0.11 -0.67  0.11  0.11  0.17  0.10  0.15  0.14  1 -6.7e-16 3.1

                       PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  PC9 PC10
SS loadings           3.15 2.16 1.02 0.84 0.78 0.60 0.47 0.38 0.31 0.29
Proportion Var        0.31 0.22 0.10 0.08 0.08 0.06 0.05 0.04 0.03 0.03
Cumulative Var        0.31 0.53 0.63 0.72 0.79 0.86 0.90 0.94 0.97 1.00
Proportion Explained  0.31 0.22 0.10 0.08 0.08 0.06 0.05 0.04 0.03 0.03
Cumulative Proportion 0.31 0.53 0.63 0.72 0.79 0.86 0.90 0.94 0.97 1.00

> # Scree plot and parallel analysis
> fa.parallel(mycor, n.obs=nobs, n.iter=100, fa="pc", nfactors=p)
Parallel analysis suggests that the number of factors =  NA  and the number of components =  2
>
```



Using Kaiser's criterion, three factors that have eigenvalues >1.0 are extracted. However, Cattell's scree plot test and parallel analysis both suggest a two-factor solution. As the third extracted factor contributes to a 10% increase in explained variance, and the 3 factors together can explain 63% variance of the variables, a three-factor solution is therefore preferred to the two-factor solution.

```
> # PC solutions with 3 factors extracted and varimax rotation
> fit_pc <- principal(mycor, n.obs=nobs, nfactors=3, residuals=TRUE, rotate="varima
> fit_pc
Principal Components Analysis
Call: principal(r = mycor, nfactors = 3, residuals = TRUE, rotate = "varimax",
    n.obs = nobs)
Standardized loadings (pattern matrix) based upon correlation matrix
      RC1   RC2   RC3   h2   u2 com
V1   0.82  0.08  0.13 0.69 0.31 1.1
V2   0.25  0.71  0.25 0.63 0.37 1.5
V3   0.65  0.03  0.12 0.44 0.56 1.1
V4   0.13  0.87  0.07 0.78 0.22 1.1
V5   0.05  0.24  0.82 0.73 0.27 1.2
V6   0.83 -0.01 -0.03 0.70 0.30 1.0
V7   0.26  0.04  0.79 0.70 0.30 1.2
V8  -0.16  0.71 -0.01 0.54 0.46 1.1
V9   0.79 -0.02  0.24 0.69 0.31 1.2
V10 -0.03  0.66  0.10 0.44 0.56 1.1

                       RC1  RC2  RC3
SS loadings           2.59 2.27 1.46
Proportion Var        0.26 0.23 0.15
Cumulative Var        0.26 0.49 0.63
Proportion Explained  0.41 0.36 0.23
Cumulative Proportion 0.41 0.77 1.00

Mean item complexity =  1.1
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is  0.09
 with the empirical chi square  66.63  with prob <  1.7e-07

Fit based upon off diagonal values = 0.9>
```

Considering the communalities based on the 3-factor solution, their range is from 0.44 (V3) to 0.78 (V4), suggesting that a three-factor solution is desirable.

More importantly, as will be shown in part (c), the three-factor solutions give a meaningful interpretation for the extracted factors.

(c) Factor rotation was conducted using Varimax. No double-loaded items were found. Factor loadings larger than 0.30 will be considered as a salient loading on a particular factor. Hence, it is found that

(i) V1 (discussing frustration), V3 (expressing emotions), V6 (seeking advice), and V9 (telling trusted people one’s feelings) load on factor 1. Factor 1 is about “help-seeking behavior”.

(ii) V2 (step-by-step plan), V4 (correcting with educational approach), V8 (taking direct action), and V10 (putting aside other activities) load on factor 2. Factor 2 is about “problem-focused coping strategies/action for correction”.

(iii) V5 (being emotionally honest with oneself) and V7 (exploring emotions caused by problem) load on factor 3. Factor 3 is about “emotion-focused coping strategies”.

(d) The 3-factor solution provides a reasonable model goodness-of-fit, because

(i) 63.3% of the variance in variables can be explained by the three-factor solution,

(ii) the communalities of the variables are moderate to high (0.44 to 0.78),

(iii) the residuals (off diagonal elements in the matrix below) are in general small,

```
> # output residuals
> error <- as.data.frame(fit_pc["residual"])
> round(error,3)
    residual.V1 residual.V2 residual.V3 residual.V4 residual.V5 residual.V6 residual.V7 residual.V8 residual.V9
V1        0.309      -0.018      -0.177      -0.036       0.020      -0.044      -0.018      -0.063      -0.119
V2       -0.018       0.368      -0.078      -0.080      -0.037       0.038      -0.014      -0.102      -0.078
V3       -0.177      -0.078       0.561       0.012      -0.050      -0.211       0.018       0.044      -0.027
V4       -0.036      -0.080       0.012       0.222      -0.037      -0.042       0.051      -0.102       0.021
V5        0.020      -0.037      -0.050      -0.037       0.273       0.075      -0.268       0.018       0.032
V6       -0.044       0.038      -0.211      -0.042       0.075       0.303       0.005       0.059      -0.101
V7       -0.018      -0.014       0.018       0.051      -0.268       0.005       0.305       0.044      -0.079
V8       -0.063      -0.102       0.044      -0.102       0.018       0.059       0.044       0.463       0.084
V9       -0.119      -0.078      -0.027       0.021       0.032      -0.101      -0.079       0.084       0.312
V10       0.096      -0.163       0.026      -0.082      -0.012      -0.059      -0.022      -0.257      -0.019
    residual.V10
V1         0.096
V2        -0.163
V3         0.026
V4        -0.082
V5        -0.012
V6        -0.059
V7        -0.022
V8        -0.257
V9        -0.019
V10        0.557
```

(iv) the extracted 3 factors can be meaningfully interpreted.