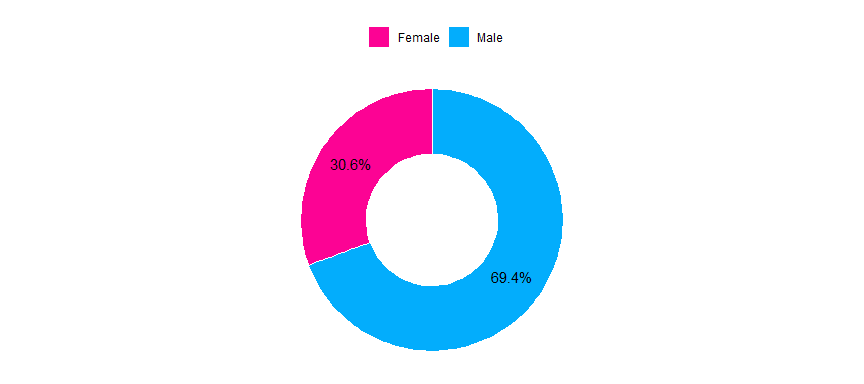
Survey Data Analysis

Ireoluwa Ige

Content

## Gender Distribution

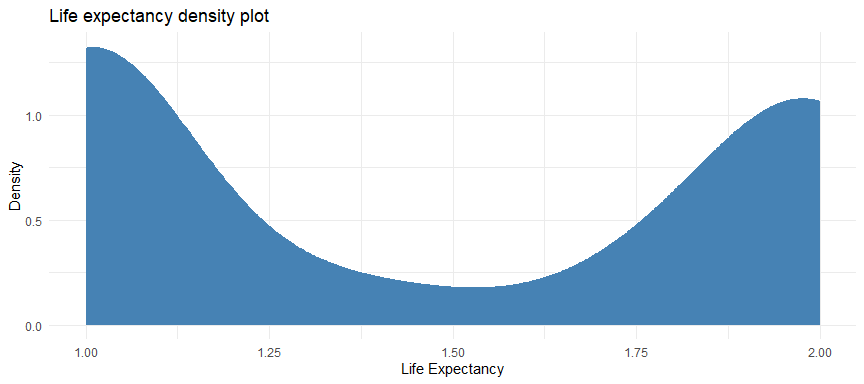


* About 31% of the participants’ were female and 69% male.

## Descriptive Statistical Analysis

### Life Expectancy

------------------------------------------------------------------------------   
Describe lf (tbl\_df, tbl, data.frame):  
  
data frame: 219 obs. of 1 variables  
 219 complete cases (100.0%)  
  
 Nr ColName Class NAs Levels  
 1 value numeric .   
  
  
------------------------------------------------------------------------------   
1 - value (numeric)  
  
 length n NAs unique 0s mean meanCI'  
 219 219 0 6 0 1.46 1.40  
 100.0% 0.0% 0.0% 1.52  
   
 .05 .10 .25 median .75 .90 .95  
 1.00 1.00 1.00 1.40 2.00 2.00 2.00  
   
 range sd vcoef mad IQR skew kurt  
 1.00 0.46 0.31 0.59 1.00 0.15 -1.85  
   
  
 value freq perc cumfreq cumperc  
1 1 98 44.7% 98 44.7%  
2 1.2 11 5.0% 109 49.8%  
3 1.4 11 5.0% 120 54.8%  
4 1.6 3 1.4% 123 56.2%  
5 1.8 21 9.6% 144 65.8%  
6 2 75 34.2% 219 100.0%  
  
' 95%-CI (classic)



* The skewness and kurtosis statistics for life expectancy is 0.15 and -1.85. By implication, the value of the skewness implies that the life expectancy is right or positively skewed such that the distribution is longer on the right side of its peak than on its left. The measure of the kurtosis (i.e. -1.85) implies that the distribution is **Platykurtic** or **negative kurtosis** as the kurtosis is less than 3. Platykurtic distributions have a low frequency of outliers.

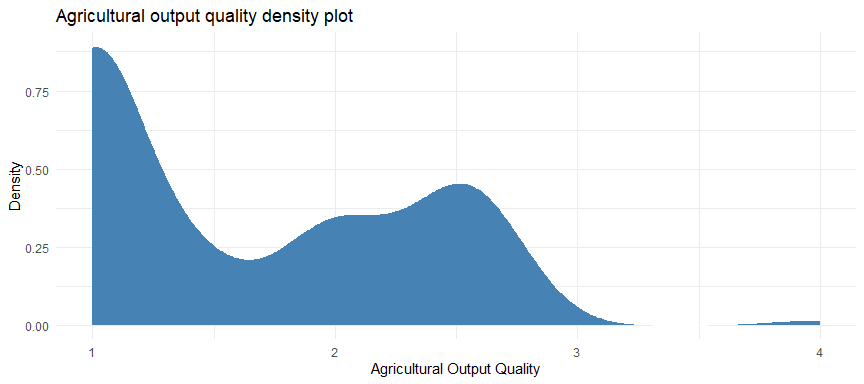
**Normality Test**

Shapiro-Wilk normality test  
  
data: .  
W = 0.72743, p-value < 2.2e-16

* The shapiro test statistic above shows that the life expectancy variable is significantly different from a normal distribution as the p-value is lower than 0.05.

### Agricultural Output Quality

------------------------------------------------------------------------------   
Describe aq (tbl\_df, tbl, data.frame):  
  
data frame: 219 obs. of 1 variables  
 219 complete cases (100.0%)  
  
 Nr ColName Class NAs Levels  
 1 value numeric .   
  
  
------------------------------------------------------------------------------   
1 - value (numeric)  
  
 length n NAs unique 0s mean meanCI'  
 219 219 0 10 0 1.64 1.55  
 100.0% 0.0% 0.0% 1.73  
   
 .05 .10 .25 median .75 .90 .95  
 1.00 1.00 1.00 1.40 2.40 2.60 2.60  
   
 range sd vcoef mad IQR skew kurt  
 3.00 0.69 0.42 0.59 1.40 0.61 -0.61  
   
  
 value freq perc cumfreq cumperc  
1 1 100 45.7% 100 45.7%  
2 1.4 20 9.1% 120 54.8%  
3 1.6 3 1.4% 123 56.2%  
4 1.8 3 1.4% 126 57.5%  
5 2 33 15.1% 159 72.6%  
6 2.2 1 0.5% 160 73.1%  
7 2.4 20 9.1% 180 82.2%  
8 2.6 36 16.4% 216 98.6%  
9 2.8 1 0.5% 217 99.1%  
10 4 2 0.9% 219 100.0%  
  
' 95%-CI (classic)



* The skewness and kurtosis statistics for agricultural output quality is 0.61 and -0.61. By implication, the value of the skewness implies that the agricultural output quality is right or positively skewed such that the distribution is longer on the right side of its peak than on its left. The measure of the kurtosis (i.e. -0.61) implies that the distribution is **Platykurtic** or **negative kurtosis** as the kurtosis is less than 3. Platykurtic distributions have a low frequency of outliers.

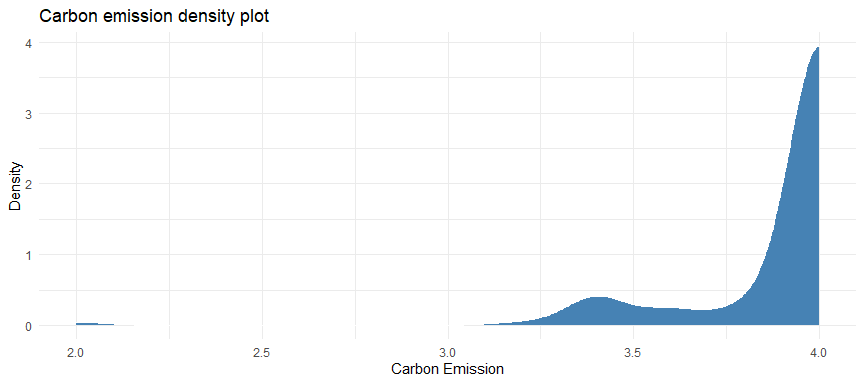
**Normality Test**

Shapiro-Wilk normality test  
  
data: .  
W = 0.80095, p-value = 5.247e-16

* The shapiro test statistic above shows that the Agricultural output quality variable is significantly different from a normal distribution as the p-value is lower than 0.05.

### Carbon Emission

------------------------------------------------------------------------------   
Describe ce (tbl\_df, tbl, data.frame):  
  
data frame: 219 obs. of 1 variables  
 219 complete cases (100.0%)  
  
 Nr ColName Class NAs Levels  
 1 value numeric .   
  
  
------------------------------------------------------------------------------   
1 - value (numeric)  
  
 length n NAs unique 0s mean meanCI'  
 219 219 0 6 0 3.90 3.86  
 100.0% 0.0% 0.0% 3.93  
   
 .05 .10 .25 median .75 .90 .95  
 3.40 3.56 4.00 4.00 4.00 4.00 4.00  
   
 range sd vcoef mad IQR skew kurt  
 2.00 0.27 0.07 0.00 0.00 -4.01 21.99  
   
  
 value freq perc cumfreq cumperc  
1 2 2 0.9% 2 0.9%  
2 3.2 2 0.9% 4 1.8%  
3 3.4 18 8.2% 22 10.0%  
4 3.6 10 4.6% 32 14.6%  
5 3.8 11 5.0% 43 19.6%  
6 4 176 80.4% 219 100.0%  
  
' 95%-CI (classic)



* The skewness and kurtosis statistics for carbon emission is -4.01 and 21.99. By implication, the value of the skewness implies that the carbon emission is left or negatively skewed such that the distribution is longer on the left side of its peak than on its right. The measure of the kurtosis (i.e. 21.99) implies that the distribution is **Leptokurtic** or **positive kurtosis** as the kurtosis is greater than 3. Leptokurtic distribution is fat-tailed, meaning that there are a lot of outliers.

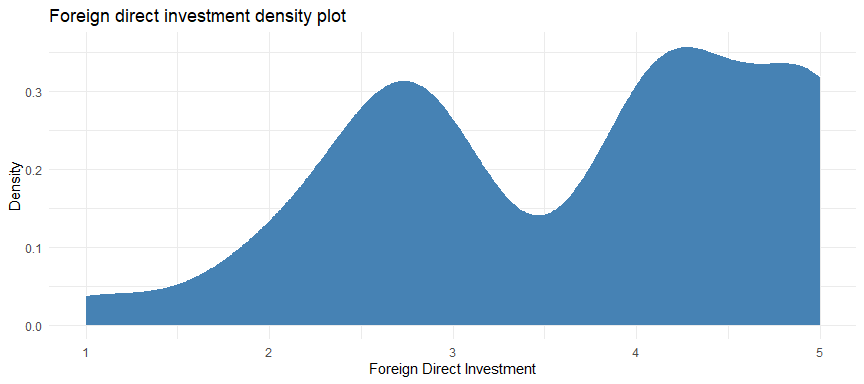
**Normality Test**

Shapiro-Wilk normality test  
  
data: .  
W = 0.43514, p-value < 2.2e-16

* The shapiro test statistic above shows that the Carbon emission variable is significantly different from a normal distribution as the p-value is lower than 0.05.

### Foreign Direct Investment

------------------------------------------------------------------------------   
Describe fd (tbl\_df, tbl, data.frame):  
  
data frame: 219 obs. of 1 variables  
 219 complete cases (100.0%)  
  
 Nr ColName Class NAs Levels  
 1 value numeric .   
  
  
------------------------------------------------------------------------------   
1 - value (numeric)  
  
 length n NAs unique 0s mean meanCI'  
 219 219 0 13 0 3.662100 3.513611  
 100.0% 0.0% 0.0% 3.810590  
   
 .05 .10 .25 median .75 .90 .95  
 2.000000 2.333333 2.666667 4.000000 4.500000 5.000000 5.000000  
   
 range sd vcoef mad IQR skew kurt  
 4.000000 1.114943 0.304455 1.482600 1.833333 -0.417893 -0.895058  
   
lowest : 1.0 (6), 1.333333, 1.666667 (3), 2.0 (11), 2.333333 (11)  
highest: 3.666667 (2), 4.0 (32), 4.333333 (37), 4.666667 (3), 5.0 (52)  
  
heap(?): remarkable frequency (23.7%) for the mode(s) (= 5)  
  
' 95%-CI (classic)



* The skewness and kurtosis statistics for foreign direct investment is -0.418 and -0.895 By implication, the value of the skewness implies that the foreign direct investment is right or positively skewed such that the distribution is longer on the right side of its peak than on its left. The measure of the kurtosis (i.e. -0.895) implies that the distribution is **Platykurtic** or **negative kurtosis** as the kurtosis is less than 3. Platykurtic distributions have a low frequency of outliers.

**Normality Test**

Shapiro-Wilk normality test  
  
data: .  
W = 0.90337, p-value = 1.072e-10

* The shapiro test statistic above shows that the Foreign direct investment variable is significantly different from a normal distribution as the p-value is lower than 0.05.

## Reliability of the scale

Some items ( CE1 CE2 CE3 CE4 CE5 ) were negatively correlated with the total scale and   
probably should be reversed.   
To do this, run the function again with the 'check.keys=TRUE' option

Reliability analysis   
Call: psych::alpha(x = questions)  
  
 raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r  
 0.81 0.83 0.95 0.21 4.9 0.018 2.6 0.35 0.042  
  
 95% confidence boundaries   
 lower alpha upper  
Feldt 0.78 0.81 0.85  
Duhachek 0.78 0.81 0.85  
  
 Reliability if an item is dropped:  
 raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r  
LFE1 0.80 0.81 0.94 0.20 4.3 0.020 0.11 0.030  
LFE2 0.80 0.81 0.94 0.20 4.3 0.020 0.11 0.030  
LFE3 0.80 0.81 0.94 0.20 4.3 0.019 0.11 0.030  
LFE4 0.80 0.81 0.94 0.20 4.3 0.019 0.11 0.030  
LFE5 0.80 0.82 0.95 0.21 4.5 0.019 0.12 0.041  
AQ1 0.79 0.81 0.94 0.20 4.3 0.020 0.11 0.040  
AQ2 0.80 0.82 0.95 0.21 4.5 0.019 0.11 0.041  
AQ3 0.78 0.81 0.94 0.20 4.2 0.021 0.11 0.035  
AQ4 0.80 0.81 0.95 0.20 4.4 0.020 0.12 0.041  
AQ5 0.78 0.81 0.94 0.20 4.2 0.021 0.11 0.030  
FD1 0.82 0.83 0.95 0.23 5.0 0.017 0.12 0.043  
FD2 0.82 0.83 0.94 0.23 5.0 0.017 0.12 0.052  
FD3 0.81 0.83 0.96 0.23 5.0 0.018 0.12 0.045  
CE1 0.82 0.83 0.95 0.23 5.1 0.018 0.12 0.067  
CE2 0.82 0.84 0.96 0.23 5.2 0.018 0.12 0.067  
CE3 0.82 0.84 0.96 0.23 5.1 0.018 0.12 0.067  
CE4 0.82 0.84 0.95 0.23 5.1 0.018 0.12 0.067  
CE5 0.82 0.83 0.95 0.23 5.0 0.018 0.12 0.054  
  
 Item statistics   
 n raw.r std.r r.cor r.drop mean sd  
LFE1 219 0.662 0.72 0.73 0.6136 1.5 0.50  
LFE2 219 0.651 0.69 0.71 0.6011 1.5 0.50  
LFE3 219 0.643 0.70 0.72 0.5916 1.5 0.50  
LFE4 219 0.631 0.68 0.69 0.5783 1.5 0.50  
LFE5 219 0.501 0.56 0.53 0.4379 1.4 0.49  
AQ1 219 0.717 0.72 0.72 0.6449 1.6 0.80  
AQ2 219 0.590 0.60 0.60 0.5192 1.5 0.62  
AQ3 219 0.759 0.77 0.78 0.6949 1.6 0.81  
AQ4 219 0.643 0.65 0.65 0.5810 1.6 0.60  
AQ5 219 0.768 0.76 0.77 0.7017 1.8 0.86  
FD1 219 0.553 0.32 0.32 0.3816 3.5 1.30  
FD2 219 0.550 0.31 0.31 0.3733 3.4 1.33  
FD3 219 0.488 0.30 0.26 0.3554 4.2 0.97  
CE1 219 0.086 0.28 0.25 0.0290 3.9 0.36  
CE2 219 0.046 0.23 0.17 0.0150 4.0 0.19  
CE3 219 0.114 0.28 0.23 0.0528 3.8 0.38  
CE4 219 0.061 0.27 0.25 0.0055 3.9 0.35  
CE5 219 0.142 0.29 0.25 0.0879 3.9 0.34  
  
Non missing response frequency for each item  
 1 2 3 4 5 miss  
LFE1 0.54 0.46 0.00 0.00 0.00 0  
LFE2 0.53 0.47 0.00 0.00 0.00 0  
LFE3 0.53 0.47 0.00 0.00 0.00 0  
LFE4 0.53 0.47 0.00 0.00 0.00 0  
LFE5 0.59 0.41 0.00 0.00 0.00 0  
AQ1 0.55 0.26 0.17 0.01 0.00 0  
AQ2 0.56 0.39 0.04 0.01 0.00 0  
AQ3 0.56 0.25 0.19 0.01 0.00 0  
AQ4 0.46 0.50 0.03 0.01 0.00 0  
AQ5 0.47 0.26 0.26 0.01 0.00 0  
FD1 0.08 0.21 0.12 0.32 0.26 0  
FD2 0.05 0.37 0.00 0.32 0.26 0  
FD3 0.04 0.05 0.02 0.49 0.40 0  
CE1 0.00 0.01 0.11 0.89 0.00 0  
CE2 0.00 0.01 0.00 0.99 0.00 0  
CE3 0.00 0.01 0.13 0.86 0.00 0  
CE4 0.00 0.01 0.10 0.89 0.00 0  
CE5 0.00 0.01 0.09 0.90 0.00 0

### How to interpret ‘Reliability if an item is dropped’?

* The overall (raw\_alpha) is 0.8132048. Each row refers to each item and has a raw alpha associated—this refers to the overall when that particular item has been dropped/deleted. For example, the first row refers to LFE1, and if it is dropped, the overall becomes 0.7962759, which reflects lower reliability, so we want to keep LFE1
* We are checking whether any of these raw alpha values are greater than the overall of 0.8132048; if yes, this means that dropping that particular item will increase the overall of the scale.
* If item CE4 is dropped, the becomes 0.8184122 which is the highest value that the could assume. Thus, we can drop the item and recalaculate the cronbach-alpha.
* The other columns of this table refer to how the other statistics will change if that particular item has been dropped/deleted.

### How to interpret the final frequency table?

This table tells us what percentage of people gave each response to each of the items (i.e., since if you have a 5-point scale, then it tells you how many percent of responses were 1, 2, 3, 4, or 5). This helps you check the distribution of responses and whether everyone is giving the same responses (which will lead to low reliability).

### Items to be dropped

* Item to be dropped is CE4

### Redo Reliability

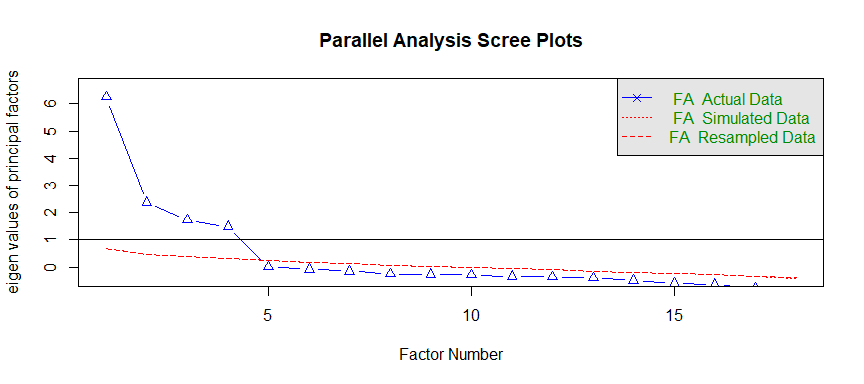
Below we redo the reliability analysis by removing the item above

Some items ( CE1 CE2 CE3 CE5 ) were negatively correlated with the total scale and   
probably should be reversed.   
To do this, run the function again with the 'check.keys=TRUE' option

* The raw alpha is now 0.8184122.

## Factor Analysis

### Parallel Analysis

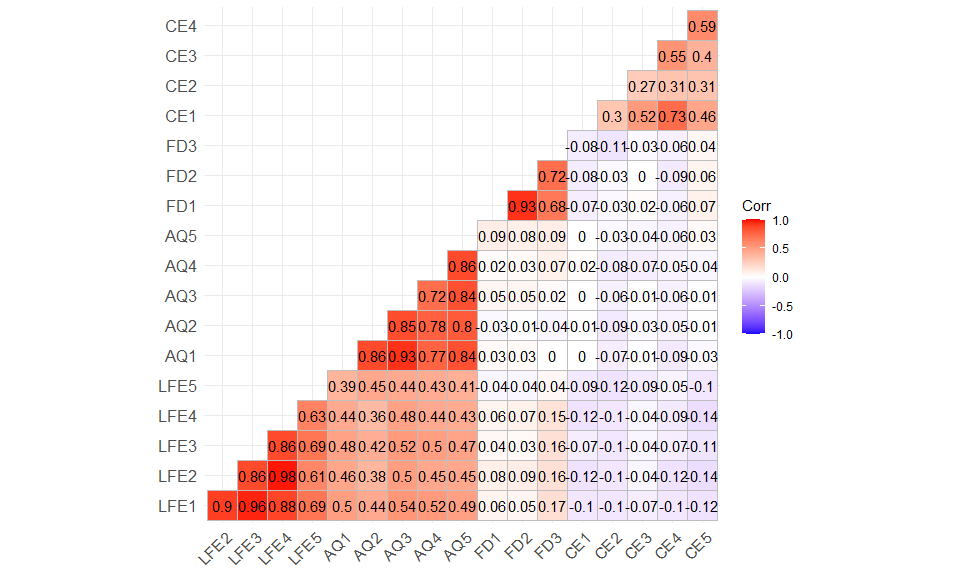


Parallel analysis suggests that the number of factors = 4 and the number of components = NA

The blue line shows eigenvalues of actual data and the red lines (placed on top of each other) show simulated and resampled data. Here we look at the large drops in the actual data and spot the point where it levels off to the right. Also we locate the point of inflection – the point where the gap between simulated data and actual data tends to be minimum.

Looking at this plot and parallel analysis, anywhere between 3 to 5 factors would be good choice.

### Correlation



### Analysis

Now that we’ve arrived at probable number number of factors, let’s start off with 4 (as the number of factors) as suggested by the scree plot above. In order to perform factor analysis, we’ll use psych package’s fa() function. Given below are the arguments we’ll supply:

* r - Raw data or correlation or covariance matrix
* nfactors – Number of factors to extract
* rotate – Although there are various types rotations, Varimax and Oblimin are most popular
* fm – One of the factor extraction techniques like Minimum Residual (OLS), Maximum Liklihood, Principal Axis etc.

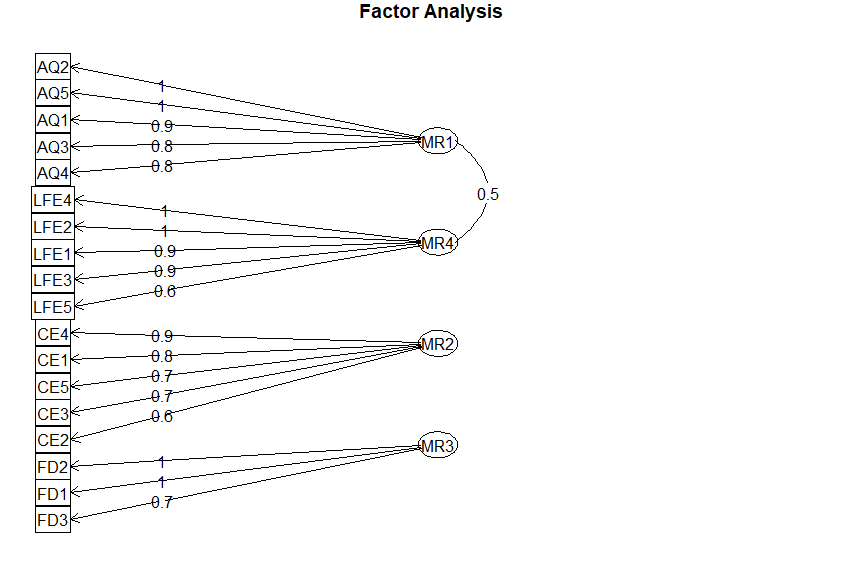
In this case, we will select oblique rotation (rotate = “oblimin”) as we believe that there is correlation in the factors. Note that Varimax rotation is used under the assumption that the factors are completely uncorrelated. We will use Ordinary Least Squared/Minres factoring (fm = “minres”), as it is known to provide results similar to Maximum Likelihood without assuming multivariate normal distribution and derives solutions through iterative eigendecomposition like principal axis.

Factor Analysis using method = minres  
Call: fa(r = questions, nfactors = 4, rotate = "oblimin", fm = "minres")  
Standardized loadings (pattern matrix) based upon correlation matrix  
 MR1 MR4 MR2 MR3 h2 u2 com  
LFE1 0.04 0.95 -0.01 -0.01 0.94 0.061 1.0  
LFE2 -0.01 0.95 -0.01 0.04 0.91 0.092 1.0  
LFE3 0.03 0.94 0.02 -0.03 0.90 0.102 1.0  
LFE4 -0.04 0.97 0.00 0.01 0.90 0.097 1.0  
LFE5 0.14 0.62 -0.04 -0.08 0.49 0.512 1.1  
AQ1 0.92 0.04 0.00 0.00 0.88 0.118 1.0  
AQ2 0.96 -0.11 -0.04 -0.09 0.83 0.168 1.1  
AQ3 0.85 0.15 0.03 0.03 0.87 0.131 1.1  
AQ4 0.83 0.03 -0.02 -0.02 0.72 0.282 1.0  
AQ5 0.96 0.01 0.03 0.09 0.93 0.068 1.0  
FD1 0.01 -0.02 0.01 0.96 0.93 0.073 1.0  
FD2 0.02 -0.03 -0.02 0.97 0.95 0.053 1.0  
FD3 -0.08 0.11 0.00 0.71 0.52 0.479 1.1  
CE1 0.06 -0.03 0.82 -0.06 0.69 0.309 1.0  
CE2 0.00 -0.02 0.63 -0.02 0.40 0.595 1.0  
CE3 -0.03 0.04 0.71 0.03 0.51 0.493 1.0  
CE4 -0.06 0.04 0.90 -0.04 0.81 0.191 1.0  
CE5 0.05 -0.08 0.74 0.12 0.57 0.434 1.1  
  
 MR1 MR4 MR2 MR3  
SS loadings 4.21 4.16 2.95 2.42  
Proportion Var 0.23 0.23 0.16 0.13  
Cumulative Var 0.23 0.46 0.63 0.76  
Proportion Explained 0.31 0.30 0.21 0.18  
Cumulative Proportion 0.31 0.61 0.82 1.00  
  
 With factor correlations of   
 MR1 MR4 MR2 MR3  
MR1 1.00 0.50 -0.06 0.02  
MR4 0.50 1.00 -0.14 0.08  
MR2 -0.06 -0.14 1.00 -0.04  
MR3 0.02 0.08 -0.04 1.00  
  
Mean item complexity = 1  
Test of the hypothesis that 4 factors are sufficient.  
  
The degrees of freedom for the null model are 153 and the objective function was 22.05 with Chi Square of 4657.18  
The degrees of freedom for the model are 87 and the objective function was 3.39   
  
The root mean square of the residuals (RMSR) is 0.02   
The df corrected root mean square of the residuals is 0.03   
  
The harmonic number of observations is 219 with the empirical chi square 30.52 with prob < 1   
The total number of observations was 219 with Likelihood Chi Square = 707.03 with prob < 7.4e-98   
  
Tucker Lewis Index of factoring reliability = 0.755  
RMSEA index = 0.18 and the 90 % confidence intervals are 0.169 0.193  
BIC = 238.18  
Fit based upon off diagonal values = 1  
Measures of factor score adequacy   
 MR1 MR4 MR2 MR3  
Correlation of (regression) scores with factors 0.99 0.99 0.95 0.99  
Multiple R square of scores with factors 0.98 0.98 0.90 0.97  
Minimum correlation of possible factor scores 0.96 0.96 0.81 0.94

Now we need to consider the loadings more than 0.5 and not loading on more than one factor. Note that negative values are acceptable here. So let’s first establish the cut off to improve visibility:

Loadings:  
 MR1 MR4 MR2 MR3   
LFE1 0.947   
LFE2 0.955   
LFE3 0.938   
LFE4 0.968   
LFE5 0.616   
AQ1 0.920   
AQ2 0.957   
AQ3 0.849   
AQ4 0.833   
AQ5 0.956   
FD1 0.964  
FD2 0.974  
FD3 0.707  
CE1 0.824   
CE2 0.631   
CE3 0.715   
CE4 0.898   
CE5 0.738   
  
 MR1 MR4 MR2 MR3  
SS loadings 4.129 4.067 2.948 2.425  
Proportion Var 0.229 0.226 0.164 0.135  
Cumulative Var 0.229 0.455 0.619 0.754

As you can see few variables have become insignificant and none of the variables have multiple-loadings. This is called simple structure



### Adequacy Test

Now that we’ve achieved a structure it’s time for us to validate our model. Let’s look at the factor analysis output to proceed:

The root mean square of residuals (RMSR) is 0.02. This is acceptable as this value should be closer to 0. Finally, the Tucker-Lewis Index (TLI) is 0.755 – an acceptable value considering it’s over 0.5.