Covid-19 and Character Strengths

# Description:

The Covid-19 pandemic obliged people around the world to stay home and self-isolate, with a number of negative psychological consequences. This study focuses on the protective role of character strengths in sustaining mental health and self-efficacy during lockdown. Data were collected from 944 Italian respondents (mean age = 37.24 years, SD = 14.50) by means of an online survey investigating character strengths, psychological distress and Covid-19-related self-efficacy one month after lockdown began. Using principal component analysis, four strengths factors were extracted, namely transcendence, interpersonal, openness and restraint. Regression models with second-order factors showed that transcendence strengths had a strong inverse association with psychological distress, and a positive association with self-efficacy. Regression models with single strengths identified hope, zest, prudence, love and forgiveness as the strengths most associated with distress, love and zest as the most related to self-efficacy and zest to general mental health. Openness factor and appreciation of beauty showed an unexpected direct relation with psychological distress. These results provide original evidence of the association of character strengths, and transcendence strengths in particular, with mental health and self-efficacy in a pandemic and are discussed within the field of positive psychology.

## Variables:

### Factors extracted through PCA

1. Openness
2. Restraint
3. Transcendence
4. Interpersonal

### The three dependent measures

1. DASS21 (Depression Anxiety and Stress Scale)
2. GHQ12 (General Health Questionnaire)
3. SEC (Self-efficacy for Covid-19)

### Six demographic variables added in the analysis:

1. Age
2. Gender
3. Work (representing the perceived work change subsequent to lockdown)
4. Student (being a student or not)
5. Day (how many days passed when the participant responded since the day the survey was opened)

## Objectives:

1. Perform Principle component analysis
2. Perform cluster analysis using the four strengths factors extracted
3. Perform Multivariate regression for the three dependent measures

## Imports

library(readxl)  
library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(summarytools)  
library(corrplot)

## corrplot 0.92 loaded

library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library(factoextra)

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library(caret)

## Loading required package: lattice

library(tidyverse)

## ── Attaching packages  
## ───────────────────────────────────────  
## tidyverse 1.3.2 ──

## ✔ tibble 3.1.8 ✔ purrr 0.3.5  
## ✔ tidyr 1.2.1 ✔ stringr 1.5.0  
## ✔ readr 2.1.3 ✔ forcats 0.5.2  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ psych::%+%() masks ggplot2::%+%()  
## ✖ psych::alpha() masks ggplot2::alpha()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ purrr::lift() masks caret::lift()  
## ✖ MASS::select() masks dplyr::select()  
## ✖ tibble::view() masks summarytools::view()

library(ROCR)  
library(MVN)  
library(car)

## Loading required package: carData  
##   
## Attaching package: 'car'  
##   
## The following object is masked from 'package:purrr':  
##   
## some  
##   
## The following object is masked from 'package:psych':  
##   
## logit  
##   
## The following object is masked from 'package:dplyr':  
##   
## recode

library(biotools)

## ---  
## biotools version 4.2

library(MVTests)

##   
## Attaching package: 'MVTests'  
##   
## The following object is masked from 'package:datasets':  
##   
## iris

DB <- read\_excel("DB.xlsx")  
DB

## # A tibble: 944 × 41  
## Participant Openn…¹ Restr…² Trans…³ Inter…⁴ DASS\_21 GHQ\_12 SEC Age Gender  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 1 102 51 116 121 15 24 11 26 Female  
## 2 2 96 49 121 100 18 15 20 25 Male   
## 3 3 99 58 120 87 0 7 20 26 Male   
## 4 4 119 68 131 84 15 18 18 25 Female  
## 5 5 127 61 151 97 37 19 14 49 Female  
## 6 6 99 43 108 123 11 14 13 22 Male   
## 7 7 127 62 113 118 30 24 14 22 Female  
## 8 8 103 53 108 121 29 13 14 31 Female  
## 9 9 96 52 109 103 3 15 16 27 Male   
## 10 10 132 64 123 98 13 18 10 27 Female  
## # … with 934 more rows, 31 more variables: Work <dbl>, Student <chr>,  
## # Day <dbl>, Sons <dbl>, Appreciation\_of\_beauty <dbl>, Bravery <dbl>,  
## # Creativity <dbl>, Curiosity <dbl>, Fairness <dbl>, Forgiveness <dbl>,  
## # Gratitude <dbl>, Honesty <dbl>, Hope <dbl>, Humilty <dbl>, Humor <dbl>,  
## # Judgment <dbl>, Kindness <dbl>, Leadership <dbl>, Love <dbl>,  
## # Love\_of\_learning <dbl>, Perseverance <dbl>, Perspective <dbl>,  
## # Prudence <dbl>, Self\_regulation <dbl>, Social\_intelligence <dbl>, …

## Check for missing values:

sum(is.na(DB))

## [1] 0

## About the dataset :

After removing the factors (previously extracted) , the 6 demographic variables ,and the DASS measures (Used in computing the DASS\_21 response variable), the total number of variables are 27, Which are then divided into 24 Explanatory and 3 response variables.Furthermore,There are no missing points .

## Descriptive statistics

DB <- DB[-c(1) ] # Dropped ID column  
DB1 <- DB[ -c(1:4,8:13,38:40) ] ## dropped factors + demographic variables  
descr(DB1)

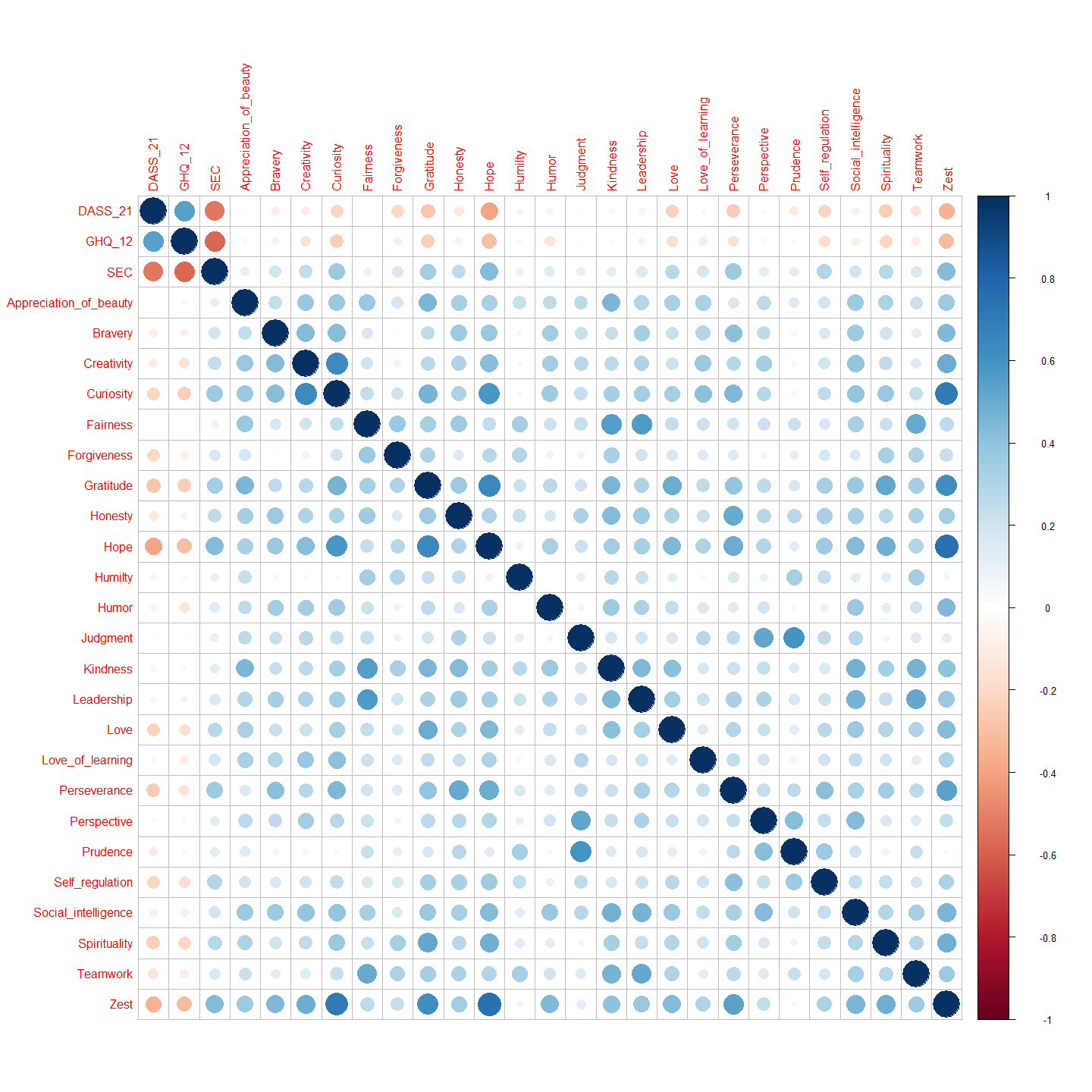
## Descriptive Statistics   
## DB1   
## N: 944   
##   
## Appreciation\_of\_beauty Bravery Creativity Curiosity DASS\_21 Fairness  
## ----------------- ------------------------ --------- ------------ ----------- --------- ----------  
## Mean 20.03 18.31 18.26 17.90 15.10 19.67  
## Std.Dev 2.89 3.20 3.41 3.30 10.70 2.64  
## Min 10.00 8.00 7.00 6.00 0.00 8.00  
## Q1 18.00 16.00 16.00 16.00 7.00 18.00  
## Median 20.00 18.00 18.00 18.00 13.00 20.00  
## Q3 22.00 20.00 20.00 20.00 20.00 21.00  
## Max 25.00 25.00 25.00 25.00 61.00 25.00  
## MAD 2.97 2.97 2.97 2.97 8.90 2.97  
## IQR 4.00 4.00 4.00 4.00 13.00 3.00  
## CV 0.14 0.17 0.19 0.18 0.71 0.13  
## Skewness -0.38 -0.21 -0.27 -0.14 1.24 -0.42  
## SE.Skewness 0.08 0.08 0.08 0.08 0.08 0.08  
## Kurtosis -0.15 0.00 -0.01 -0.11 1.69 0.44  
## N.Valid 944.00 944.00 944.00 944.00 944.00 944.00  
## Pct.Valid 100.00 100.00 100.00 100.00 100.00 100.00  
##   
## Table: Table continues below  
##   
##   
##   
## Forgiveness GHQ\_12 Gratitude Honesty Hope Humilty Humor  
## ----------------- ------------- -------- ----------- --------- -------- --------- --------  
## Mean 17.36 17.01 18.88 21.31 17.50 17.56 18.63  
## Std.Dev 3.76 4.90 3.35 2.33 3.56 3.25 3.34  
## Min 5.00 3.00 6.00 12.00 5.00 6.00 10.00  
## Q1 15.00 14.00 17.00 20.00 15.00 15.00 16.00  
## Median 18.00 17.00 19.00 21.00 18.00 18.00 19.00  
## Q3 20.00 20.00 21.00 23.00 20.00 20.00 21.00  
## Max 25.00 31.00 25.00 25.00 25.00 25.00 25.00  
## MAD 4.45 4.45 2.97 2.97 2.97 2.97 2.97  
## IQR 5.00 6.00 4.00 3.00 5.00 5.00 5.00  
## CV 0.22 0.29 0.18 0.11 0.20 0.18 0.18  
## Skewness -0.36 0.08 -0.36 -0.57 -0.48 -0.24 -0.24  
## SE.Skewness 0.08 0.08 0.08 0.08 0.08 0.08 0.08  
## Kurtosis -0.05 0.12 -0.02 0.51 0.28 -0.05 -0.46  
## N.Valid 944.00 944.00 944.00 944.00 944.00 944.00 944.00  
## Pct.Valid 100.00 100.00 100.00 100.00 100.00 100.00 100.00  
##   
## Table: Table continues below  
##   
##   
##   
## Judgment Kindness Leadership Love Love\_of\_learning Perseverance  
## ----------------- ---------- ---------- ------------ -------- ------------------ --------------  
## Mean 20.48 20.65 17.99 19.38 17.58 18.88  
## Std.Dev 2.65 2.58 3.04 3.28 3.61 3.41  
## Min 9.00 11.00 8.00 7.00 7.00 5.00  
## Q1 19.00 19.00 16.00 17.00 15.00 17.00  
## Median 20.00 21.00 18.00 20.00 17.00 19.00  
## Q3 22.00 23.00 20.00 22.00 20.00 21.00  
## Max 25.00 25.00 25.00 25.00 25.00 25.00  
## MAD 2.97 2.97 2.97 2.97 4.45 2.97  
## IQR 3.00 4.00 4.00 5.00 5.00 4.00  
## CV 0.13 0.13 0.17 0.17 0.21 0.18  
## Skewness -0.48 -0.49 -0.07 -0.54 -0.01 -0.57  
## SE.Skewness 0.08 0.08 0.08 0.08 0.08 0.08  
## Kurtosis 0.41 0.20 -0.24 0.09 -0.65 0.38  
## N.Valid 944.00 944.00 944.00 944.00 944.00 944.00  
## Pct.Valid 100.00 100.00 100.00 100.00 100.00 100.00  
##   
## Table: Table continues below  
##   
##   
##   
## Perspective Prudence SEC Self\_regulation Social\_intelligence  
## ----------------- ------------- ---------- -------- ----------------- ---------------------  
## Mean 17.75 17.71 15.06 16.77 18.63  
## Std.Dev 3.20 3.29 3.99 3.52 2.70  
## Min 6.00 8.00 5.00 6.00 7.00  
## Q1 16.00 16.00 12.00 14.50 17.00  
## Median 18.00 18.00 15.00 17.00 19.00  
## Q3 20.00 20.00 18.00 19.00 20.00  
## Max 25.00 25.00 25.00 25.00 25.00  
## MAD 2.97 2.97 4.45 2.97 2.97  
## IQR 4.00 4.00 6.00 4.25 3.00  
## CV 0.18 0.19 0.26 0.21 0.15  
## Skewness -0.12 -0.24 0.07 -0.28 -0.23  
## SE.Skewness 0.08 0.08 0.08 0.08 0.08  
## Kurtosis 0.09 -0.24 -0.10 -0.30 0.13  
## N.Valid 944.00 944.00 944.00 944.00 944.00  
## Pct.Valid 100.00 100.00 100.00 100.00 100.00  
##   
## Table: Table continues below  
##   
##   
##   
## Spirituality Teamwork Zest  
## ----------------- -------------- ---------- --------  
## Mean 15.29 18.31 17.27  
## Std.Dev 4.09 2.70 3.54  
## Min 5.00 8.00 6.00  
## Q1 12.00 17.00 15.00  
## Median 15.00 18.00 17.00  
## Q3 18.00 20.00 20.00  
## Max 25.00 25.00 25.00  
## MAD 4.45 2.97 4.45  
## IQR 6.00 3.00 5.00  
## CV 0.27 0.15 0.21  
## Skewness 0.04 -0.27 -0.18  
## SE.Skewness 0.08 0.08 0.08  
## Kurtosis -0.47 0.27 -0.30  
## N.Valid 944.00 944.00 944.00  
## Pct.Valid 100.00 100.00 100.00

### Testing for the correlations between the variables

Heatmap

Barlett’s test which is used ro determine if factor analysis is appropriate.

DB\_corr <- cor(DB1)  
corrplot(DB\_corr, method="circle")



From the correlogram above , it is observed that the there are high correlations between our dependent variables which lay the framework for why Multivariate regression was chosen as the analysis method . Moreover, there appears to be high correlations between the explanatory variables , which will violate the multicolinearity assumption later on the Modelling section . Principal component Analysis will not only solve the multicolinearity issue at hand but also help in data reduction.

cortest.bartlett(DB1, n= nrow(DB1))

## R was not square, finding R from data

## $chisq  
## [1] 11225.26  
##   
## $p.value  
## [1] 0  
##   
## $df  
## [1] 351

Since the P-value is smaller than 0.05 , it is safe to proceed with PCA and FA.

# Data Reduction:

Considering the multitude of explanatory variables , a data reduction technique is advised. Two common methods for data reduction are the Prinicpal component analysis and the factor analysis. Each has their underlying assumptions , with some significant overlap. However, PCA is often used when the objective is prediction rather than interpretation. Thus, factor analysis is recommended as interpreting the results is the main objective.

## Principal component analysis:

Principal component analysis (PCA) is a data reduction technique .It aims at transforming a set of interrelated variables to a smaller set of uncorrelated variables.PCs are a linear combination of the variables, such that:

PC\_data <- DB1[ -c(1:3)] # dropping the Response (dependent) variables  
PC <- prcomp(PC\_data, scale = TRUE) # Scaling is equivalent to using correlation matrix R  
summary(PC)

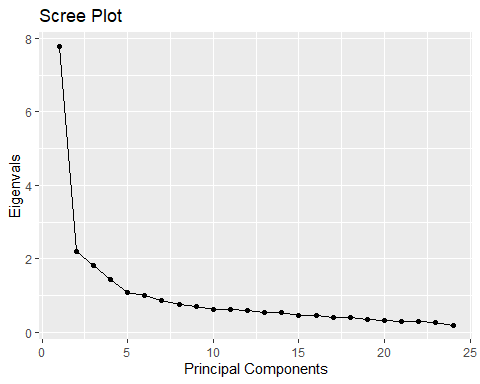
## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6 PC7  
## Standard deviation 2.7891 1.48118 1.35179 1.20365 1.04435 0.99740 0.93171  
## Proportion of Variance 0.3241 0.09141 0.07614 0.06037 0.04544 0.04145 0.03617  
## Cumulative Proportion 0.3241 0.41553 0.49167 0.55204 0.59748 0.63893 0.67510  
## PC8 PC9 PC10 PC11 PC12 PC13 PC14  
## Standard deviation 0.87452 0.8358 0.7960 0.78731 0.77021 0.73259 0.72611  
## Proportion of Variance 0.03187 0.0291 0.0264 0.02583 0.02472 0.02236 0.02197  
## Cumulative Proportion 0.70697 0.7361 0.7625 0.78830 0.81302 0.83538 0.85735  
## PC15 PC16 PC17 PC18 PC19 PC20 PC21  
## Standard deviation 0.68685 0.66953 0.6425 0.62773 0.59554 0.56257 0.53487  
## Proportion of Variance 0.01966 0.01868 0.0172 0.01642 0.01478 0.01319 0.01192  
## Cumulative Proportion 0.87700 0.89568 0.9129 0.92930 0.94408 0.95726 0.96918  
## PC22 PC23 PC24  
## Standard deviation 0.53073 0.52156 0.43114  
## Proportion of Variance 0.01174 0.01133 0.00775  
## Cumulative Proportion 0.98092 0.99225 1.00000

In almost all social sciences , a 50% or more explained variance is acceptable . Thus, taking the first 4 PCs will satisfy that requirement. Moreover , plotting a scree plot will aid in determining the numbers of PCs to be taken .

### Scree plot

Eigenvals = PC$sdev ^ 2  
  
qplot(c(1:24), Eigenvals) +   
 geom\_line() +   
 xlab("Principal Components") +   
 ggtitle("Scree Plot")

## Warning: `qplot()` was deprecated in ggplot2 3.4.0.



Judging by the scree plot above and the PCA summary it safe to conclude that taking 4 PCs is sufficient but not necessary. As mentioned in the beginning of the data reduction chapter , the number of observations are greater than the 20p rule of thumb .However, PCA will lay the framework when factors Analysis is applied .

## Factor Analysis

After extracting the 4 components , factor analysis is used in order to uncover patterns within the data. Although PCA and factor analysis are somewhat similar, it is important to point out that factor analysis presumes that the observed correlations among the variables reflect a specific underlying structure.Moreover , Factors obtained through factor analysis can be interpreted as real life unobserved variables.Taking into considerations the the theoretical background of the factors and the variables themselves , it can be assumed that the factors will be correlated . Thus, an oblique (promax) rotation would be appropriate.

set.seed(120)  
covid\_FA<- factanal(PC\_data, factors = 4 ,rotation = "promax", scores = "Bartlett")  
covid\_FA$loadings

##   
## Loadings:  
## Factor1 Factor2 Factor3 Factor4  
## Appreciation\_of\_beauty 0.376 0.242   
## Bravery 0.596   
## Creativity 0.709   
## Curiosity -0.101 0.448 0.622   
## Fairness 0.900 -0.225   
## Forgiveness 0.383 0.246 -0.189   
## Gratitude 0.118 0.666   
## Honesty 0.299 0.157 0.220   
## Hope -0.174 0.816 0.282   
## Humilty 0.489 -0.394 0.205   
## Humor 0.249 0.423 -0.212   
## Judgment -0.181 0.205 0.797   
## Kindness 0.776 -0.105   
## Leadership 0.657 -0.106 0.211   
## Love 0.205 0.369   
## Love\_of\_learning 0.391   
## Perseverance -0.139 0.529 0.182 0.207   
## Perspective 0.302 0.532   
## Prudence -0.262 0.900   
## Self\_regulation 0.400 0.349   
## Social\_intelligence 0.330 0.345   
## Spirituality 0.580   
## Teamwork 0.664 -0.103   
## Zest 0.770 0.428 -0.175   
##   
## Factor1 Factor2 Factor3 Factor4  
## SS loadings 3.213 3.026 2.549 2.109  
## Proportion Var 0.134 0.126 0.106 0.088  
## Cumulative Var 0.134 0.260 0.366 0.454

DB1 <- cbind(DB,covid\_FA$scores)

Finally, the 4 factors extracted are as follows :

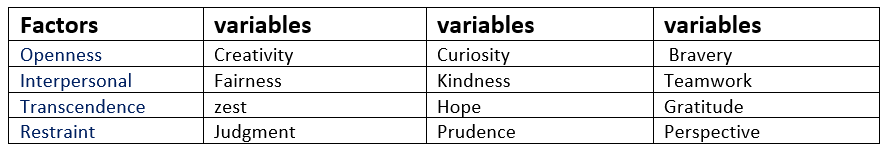


Figure 1: factors

**Openness** is how open-minded, imaginative, creative and insightful a person is or can be. As reflected by the *Creativity*, *Curiosity*,and *Bravery* variables.

***Interpersonal*** skills are the skills required to effectively communicate, interact, and work with individuals and groups. Which is reflected by the *Fairness* , *Kindness* , and *Teamwork* variables.

***Transcendence*** is an inherent human personality *trait* relating to the experience of spiritual aspects of the self. As such its reflected by the *Zest* , *Hope ,* and *Gratitude* variables.

**Restraint** is reflected by the *Judgement* , *Prudence* , and *Perspective* variables

Since , the factors extracted through our analysis match the Original factors (loadings wise) , the rest of the analysis will be conducted using the original Factors .

L2 <- covid\_FA$loadings^2  
 com <- apply(L2,1, FUN = sum)  
 ComTable <- data.frame(com)  
 colnames(ComTable) <- c("Communalities")  
 ComTable

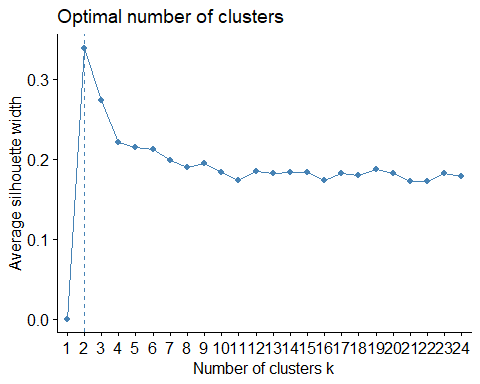
## Communalities  
## Appreciation\_of\_beauty 0.2041592  
## Bravery 0.3677879  
## Creativity 0.5138071  
## Curiosity 0.6003210  
## Fairness 0.8600448  
## Forgiveness 0.2460219  
## Gratitude 0.4616125  
## Honesty 0.1697631  
## Hope 0.7764872  
## Humilty 0.4361705  
## Humor 0.2914745  
## Judgment 0.7113967  
## Kindness 0.6229763  
## Leadership 0.4869113  
## Love 0.1884625  
## Love\_of\_learning 0.1688251  
## Perseverance 0.3752063  
## Perspective 0.3795514  
## Prudence 0.8782459  
## Self\_regulation 0.2922058  
## Social\_intelligence 0.2419256  
## Spirituality 0.3534518  
## Teamwork 0.4571545  
## Zest 0.8134679

Each value in the table corresponds to how much all 4 factors explain the variations in the variables.

## Cluster Analysis : K-means

### Silhouette statistic

FA\_OG <- DB[,1:4]  
fviz\_nbclust(FA\_OG, FUNcluster=kmeans, k.max = 24)

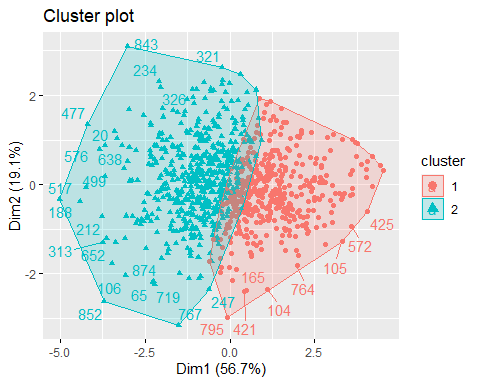


As seen in the figure above, the optimal number of clusters is 2 . Moreover, this fits our prediction about the number of clusters within the data . The data itself is concerned with the self efficacy during covid and the mental state of the subjects , thus it would make sense to have two clusters : those who had a better mental state subsequently performing well in self efficacy and those who had a worse mental state thus falling into depression and anxiety and performing poorly in self efficacy .

set.seed(120) # to allow for the reproducibility of the analysis  
DB\_cluster <- kmeans(DB[,1:4], 2, nstart = 25) # clustering on the factors already present in the data set   
DB\_cluster

## K-means clustering with 2 clusters of sizes 412, 532  
##   
## Cluster means:  
## Openness Restraint Transcendence Interpersonal  
## 1 119.3398 58.68204 138.7864 118.9733  
## 2 101.5414 53.81015 112.4699 105.7782  
##   
## Clustering vector:  
## [1] 2 2 2 2 1 2 1 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 1 1 1 1 1 2 2 2 2 1 2 2 2 1  
## [38] 2 2 1 1 2 2 2 1 2 2 2 2 2 1 1 1 2 2 1 2 1 2 2 1 1 2 2 2 1 2 2 2 2 1 2 1 2  
## [75] 1 2 1 1 1 1 1 1 1 1 2 2 2 1 1 2 2 1 2 1 1 2 1 2 2 2 1 1 1 1 1 2 2 1 1 1 2  
## [112] 2 2 2 1 1 2 2 1 2 2 2 1 2 1 2 1 2 1 2 2 1 2 2 2 2 2 1 1 1 2 2 2 1 1 1 1 2  
## [149] 2 2 1 2 1 2 1 1 1 2 2 1 2 2 2 1 1 2 1 1 2 1 2 1 2 2 1 1 1 1 1 2 2 2 1 2 2  
## [186] 2 1 2 2 1 1 1 2 1 2 1 1 2 2 2 2 1 1 2 2 1 1 2 1 1 1 2 2 2 1 2 2 1 1 2 2 1  
## [223] 1 2 2 2 1 1 2 1 2 2 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1 2 2 2 1 1 2 1 1 1  
## [260] 1 1 1 1 2 1 1 2 2 1 1 2 2 2 1 1 2 2 1 1 1 1 2 1 1 1 1 1 2 2 2 2 2 2 1 1 2  
## [297] 2 2 2 1 2 2 2 1 2 2 2 2 1 2 1 2 2 1 1 1 1 2 2 2 2 2 2 1 2 2 1 2 2 2 2 1 1  
## [334] 1 2 2 1 1 1 1 2 1 2 2 2 1 1 1 1 2 2 1 2 2 2 2 1 2 2 1 2 1 2 2 2 2 2 1 2 2  
## [371] 2 2 2 1 1 1 1 2 1 2 1 1 2 2 1 2 2 2 1 2 1 1 1 2 1 1 2 2 1 1 2 2 1 1 1 1 2  
## [408] 2 1 1 1 1 2 2 1 1 1 1 2 2 1 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 2 2 1 1 2 1 2  
## [445] 1 2 2 2 2 1 1 1 2 1 2 2 2 1 2 1 1 1 2 1 2 2 1 1 1 1 1 2 2 2 2 2 2 1 2 2 1  
## [482] 2 1 1 2 2 2 1 2 1 1 2 1 2 1 1 2 2 2 2 1 1 1 2 1 2 1 2 2 1 1 2 1 2 2 2 2 2  
## [519] 1 2 2 2 1 1 2 1 2 1 1 2 2 1 2 1 1 1 2 2 2 2 1 2 2 2 1 2 1 2 1 2 1 2 2 1 2  
## [556] 1 2 1 2 1 2 2 2 2 2 2 2 2 1 2 2 1 1 1 2 2 1 2 2 1 1 2 2 1 1 2 2 2 1 2 1 2  
## [593] 1 1 2 2 2 1 2 2 1 2 2 1 2 2 2 1 1 2 2 2 2 2 2 2 1 2 1 2 2 1 2 2 2 2 2 1 2  
## [630] 2 2 1 2 1 1 2 2 2 1 1 2 1 1 1 1 2 1 2 2 1 1 2 1 2 1 2 2 1 1 2 1 1 1 2 2 2  
## [667] 1 1 2 1 1 2 2 2 1 1 2 2 2 2 2 1 1 2 1 2 2 2 2 2 1 1 1 2 2 1 1 1 1 2 2 1 1  
## [704] 2 1 2 1 1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 1 1 2  
## [741] 2 1 2 1 2 2 1 1 2 1 1 1 2 2 2 2 2 1 2 2 1 1 1 1 2 1 2 2 1 2 1 2 2 2 1 1 2  
## [778] 2 2 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 1 1 1 1 2 2 2 2 1 2 1 2 2 1 1 2 2 1 1  
## [815] 1 2 2 1 2 1 1 2 1 2 2 1 2 1 1 1 2 2 1 2 2 2 2 2 1 2 1 1 2 1 2 2 2 2 2 2 2  
## [852] 2 2 1 1 2 1 2 2 2 2 1 2 2 1 2 1 1 1 1 1 2 2 2 2 1 1 2 1 1 2 1 2 2 2 1 1 2  
## [889] 1 2 1 1 1 1 2 2 2 2 1 2 1 2 1 1 1 1 2 1 2 2 2 2 1 2 1 2 1 1 2 1 2 1 2 2 1  
## [926] 2 1 1 2 2 1 1 2 1 1 1 2 2 1 1 1 1 2 2  
##   
## Within cluster sum of squares by cluster:  
## [1] 149555.7 238376.3  
## (between\_SS / total\_SS = 41.9 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
## [6] "betweenss" "size" "iter" "ifault"

#plot   
fviz\_cluster(DB\_cluster, data = DB[,1:4], repel=TRUE)



As seen in the plot above , The data does indeed reflect 2 clusters . Confirming ,yet again , our hypothesis of the factor correlations.

### Discriminant Analysis

Linear discriminant analysis (LDA) must be applied in order to validate our clusters accuracy.  
LDA aims at maximizing the distance between clusters in other words, minimizing the overlap between them .

### Assumptions

1. Multivariate normality
2. Equality of covariances
3. Unequal variable means

##### Equality of co-variances

factors <- DB[,1:4]  
Bo1 <- boxM(factors,DB\_cluster$cluster)  
Bo1

##   
## Box's M-test for Homogeneity of Covariance Matrices  
##   
## data: factors  
## Chi-Sq (approx.) = 83.138, df = 10, p-value = 1.215e-13

Thus the assumption of homogeneity of covariance matrices is violated.

##### Unequal Variable means

testm <- TwoSamplesHT2(factors, DB\_cluster$cluster)  
testm$p.value

## [,1]  
## [1,] 1.290117e-197

Unequal variable means assumption is satisfied

#### Modeling and Testing

DB1 <- cbind(factors ,DB\_cluster$cluster)  
covid\_DA <- qda(DB\_cluster$cluster ~ Openness + Transcendence + Restraint + Interpersonal,data = DB1, CV = T)  
  
yhat <- covid\_DA$class  
ytrue <- DB\_cluster$cluster  
table(ytrue, yhat, dnn = c('Actual Group','Predicted Group'))

## Predicted Group  
## Actual Group 1 2  
## 1 385 27  
## 2 9 523

CM <- table(list(predicted=covid\_DA$class, observed=DB\_cluster$cluster)) #confusion matrix  
caret::confusionMatrix(CM)

## Confusion Matrix and Statistics  
##   
## observed  
## predicted 1 2  
## 1 385 9  
## 2 27 523  
##   
## Accuracy : 0.9619   
## 95% CI : (0.9476, 0.9731)  
## No Information Rate : 0.5636   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.9221   
##   
## Mcnemar's Test P-Value : 0.004607   
##   
## Sensitivity : 0.9345   
## Specificity : 0.9831   
## Pos Pred Value : 0.9772   
## Neg Pred Value : 0.9509   
## Prevalence : 0.4364   
## Detection Rate : 0.4078   
## Detection Prevalence : 0.4174   
## Balanced Accuracy : 0.9588   
##   
## 'Positive' Class : 1   
##

As observed above , The LDA model has a 96.2% accuracy . False positive rate at 2.28%

## Multivariate Regression

Response variables:

1. DASS\_21 (Depression Anxiety and Stress Scale)
2. GHQ12 (General Health Questionnaire)
3. SEC (Self-efficacy for Covid-19)

Explanatory variables:

1. Openness
2. Restraint
3. Transcendence
4. Interpersonal
5. Age
6. Gender
7. Work
8. Student
9. Day

## Modelling

### Encoding the nominal variables

DB$Gender <- ifelse(DB$Gender == "Male",1,0) # encoding non numeric variable #male = 1  
DB$Student <- ifelse(DB$Student == "Student",1,0) #student = 1

### Testing for Normality

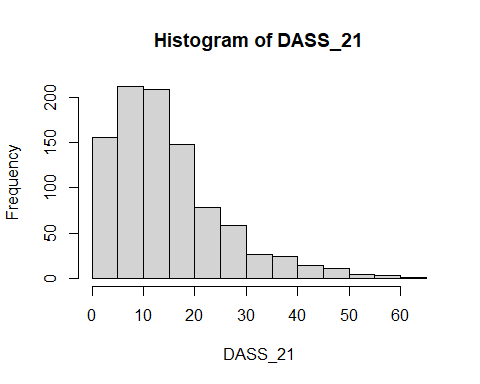
Ntest <- mvn(data = DB[,5:7], mvnTest = "mardia")  
Ntest

## $multivariateNormality  
## Test Statistic p value Result  
## 1 Mardia Skewness 325.331143087038 6.77572908779562e-64 NO  
## 2 Mardia Kurtosis 10.0459011903815 0 NO  
## 3 MVN <NA> <NA> NO  
##   
## $univariateNormality  
## Test Variable Statistic p value Normality  
## 1 Anderson-Darling DASS\_21 21.6499 <0.001 NO   
## 2 Anderson-Darling GHQ\_12 3.3987 <0.001 NO   
## 3 Anderson-Darling SEC 3.3497 <0.001 NO   
##   
## $Descriptives  
## n Mean Std.Dev Median Min Max 25th 75th Skew Kurtosis  
## DASS\_21 944 15.10064 10.697788 13 0 61 7 20 1.24192535 1.6851585  
## GHQ\_12 944 17.01059 4.903944 17 3 31 14 20 0.08351192 0.1174579  
## SEC 944 15.05508 3.987139 15 5 25 12 18 0.06906467 -0.1008443

DASS\_21<-as.matrix(DB[5])  
GHQ\_12<-as.matrix(DB[6])  
SEC<- as.matrix(DB[7])

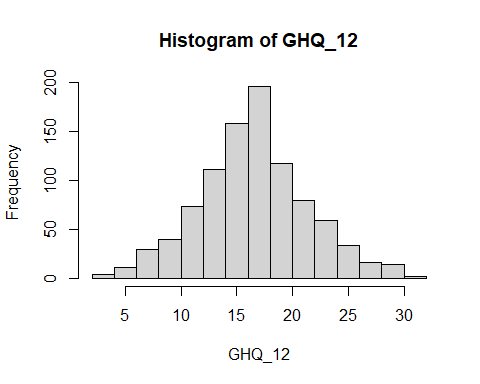
#### Histograms

hist(DASS\_21)

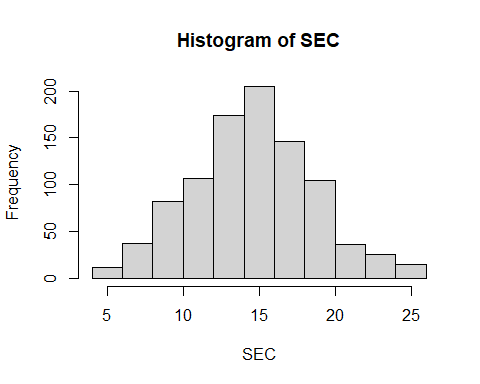


As seen the histogram above , DASS\_21 is not normally distributed.

hist(GHQ\_12)

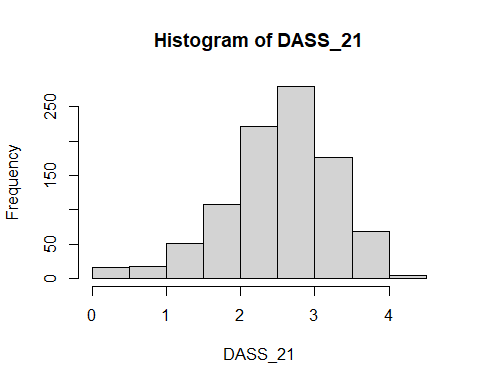


hist(SEC)



#### Transforming the DASS\_21 variable

DB[,5] <- DB[,5] + 1 # to avoid zero values, while maintaining the same distribution  
DB[,5] <- log(DB[,5])  
DASS\_21 <- as.matrix(DB[,5])  
hist(DASS\_21)



#### Fitting

mlm1 <- lm(cbind(DASS\_21, GHQ\_12 , SEC ) ~ Openness + Restraint + Transcendence +Interpersonal + Age + Gender + Work + Student + Day, data = DB)  
summary(mlm1)

## Response DASS\_21 :  
##   
## Call:  
## lm(formula = DASS\_21 ~ Openness + Restraint + Transcendence +   
## Interpersonal + Age + Gender + Work + Student + Day, data = DB)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.6043 -0.3830 0.1099 0.4616 1.8912   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.0307749 0.2677088 15.057 < 2e-16 \*\*\*  
## Openness 0.0058111 0.0021862 2.658 0.00799 \*\*   
## Restraint -0.0002008 0.0033498 -0.060 0.95220   
## Transcendence -0.0181545 0.0018225 -9.961 < 2e-16 \*\*\*  
## Interpersonal 0.0019964 0.0021755 0.918 0.35903   
## Age -0.0036803 0.0019256 -1.911 0.05628 .   
## Gender -0.2328689 0.0525125 -4.435 1.03e-05 \*\*\*  
## Work 0.0461671 0.0147651 3.127 0.00182 \*\*   
## Student 0.1124314 0.0654616 1.718 0.08622 .   
## Day -0.0070111 0.0037161 -1.887 0.05951 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6879 on 934 degrees of freedom  
## Multiple R-squared: 0.1869, Adjusted R-squared: 0.1791   
## F-statistic: 23.86 on 9 and 934 DF, p-value: < 2.2e-16  
##   
##   
## Response GHQ\_12 :  
##   
## Call:  
## lm(formula = GHQ\_12 ~ Openness + Restraint + Transcendence +   
## Interpersonal + Age + Gender + Work + Student + Day, data = DB)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.1372 -2.8826 0.0653 2.9094 13.4852   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 23.422451 1.790780 13.079 < 2e-16 \*\*\*  
## Openness -0.008209 0.014624 -0.561 0.57471   
## Restraint 0.043995 0.022407 1.963 0.04990 \*   
## Transcendence -0.104390 0.012191 -8.563 < 2e-16 \*\*\*  
## Interpersonal 0.030859 0.014553 2.121 0.03423 \*   
## Age 0.031909 0.012881 2.477 0.01342 \*   
## Gender -1.004019 0.351271 -2.858 0.00435 \*\*   
## Work 0.177025 0.098768 1.792 0.07340 .   
## Student 1.117689 0.437891 2.552 0.01086 \*   
## Day -0.024108 0.024858 -0.970 0.33239   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.601 on 934 degrees of freedom  
## Multiple R-squared: 0.128, Adjusted R-squared: 0.1196   
## F-statistic: 15.23 on 9 and 934 DF, p-value: < 2.2e-16  
##   
##   
## Response SEC :  
##   
## Call:  
## lm(formula = SEC ~ Openness + Restraint + Transcendence + Interpersonal +   
## Age + Gender + Work + Student + Day, data = DB)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.8143 -2.0455 -0.0615 2.2456 12.8834   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.1849130 1.3412492 1.629 0.103647   
## Openness 0.0124604 0.0109531 1.138 0.255570   
## Restraint 0.0000374 0.0167826 0.002 0.998222   
## Transcendence 0.1086310 0.0091311 11.897 < 2e-16 \*\*\*  
## Interpersonal -0.0197282 0.0108995 -1.810 0.070614 .   
## Age 0.0062761 0.0096476 0.651 0.515508   
## Gender 1.0156289 0.2630931 3.860 0.000121 \*\*\*  
## Work -0.1476472 0.0739748 -1.996 0.046234 \*   
## Student -0.4038437 0.3279693 -1.231 0.218503   
## Day 0.0302928 0.0186181 1.627 0.104061   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.446 on 934 degrees of freedom  
## Multiple R-squared: 0.26, Adjusted R-squared: 0.2529   
## F-statistic: 36.47 on 9 and 934 DF, p-value: < 2.2e-16

The output shows simply three uni-variate multiple regression model fitted to our dependent variables , it is not sufficient to determine which variable are significant from the output above.

### Feature Selection

#### MANOVA

To determine which variables are jointly significant MANOVA has to be carried out

Manova(mlm1)

##   
## Type II MANOVA Tests: Pillai test statistic  
## Df test stat approx F num Df den Df Pr(>F)   
## Openness 1 0.013869 4.369 3 932 0.0045867 \*\*   
## Restraint 1 0.006097 1.906 3 932 0.1269813   
## Transcendence 1 0.161593 59.877 3 932 < 2.2e-16 \*\*\*  
## Interpersonal 1 0.005603 1.750 3 932 0.1550763   
## Age 1 0.020097 6.372 3 932 0.0002825 \*\*\*  
## Gender 1 0.026612 8.493 3 932 1.435e-05 \*\*\*  
## Work 1 0.011290 3.547 3 932 0.0141790 \*   
## Student 1 0.007510 2.351 3 932 0.0709526 .   
## Day 1 0.004946 1.544 3 932 0.2014911   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Restraint , Interpersonal , Student , and Day variables are all jointly insignificant.Thus, we need to update our model without these variables. Testing to check whether dropping these variable have a significant impact on the model or not .

mlm2 <- update(mlm1, . ~ . - Interpersonal - Day )  
anova(mlm1, mlm2)

## Analysis of Variance Table  
##   
## Model 1: cbind(DASS\_21, GHQ\_12, SEC) ~ Openness + Restraint + Transcendence +   
## Interpersonal + Age + Gender + Work + Student + Day  
## Model 2: cbind(DASS\_21, GHQ\_12, SEC) ~ Openness + Restraint + Transcendence +   
## Age + Gender + Work + Student  
## Res.Df Df Gen.var. Pillai approx F num Df den Df Pr(>F)  
## 1 934 4.1379   
## 2 936 2 4.1439 0.010776 1.6848 6 1866 0.1208

Judging by the anova above where the significance of two the 2 models where compared , simple model seems to explain just as much variations in the dependent variables as the complex model .If the resulting p-value is sufficiently low (usually less than 0.05), we conclude that the more complex model is significantly better than the simpler model, and thus favor the more complex model. If the p-value is not sufficiently low (usually greater than 0.05), we should favor the simpler model.\* (source: Bookdown.org)

summary(mlm2)

## Response DASS\_21 :  
##   
## Call:  
## lm(formula = DASS\_21 ~ Openness + Restraint + Transcendence +   
## Age + Gender + Work + Student, data = DB)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5965 -0.3801 0.1063 0.4647 1.7775   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.0519931 0.2379019 17.032 < 2e-16 \*\*\*  
## Openness 0.0058625 0.0021826 2.686 0.00736 \*\*   
## Restraint 0.0006899 0.0032864 0.210 0.83376   
## Transcendence -0.0176208 0.0017371 -10.144 < 2e-16 \*\*\*  
## Age -0.0039831 0.0019072 -2.088 0.03702 \*   
## Gender -0.2325733 0.0521483 -4.460 9.2e-06 \*\*\*  
## Work 0.0468828 0.0147797 3.172 0.00156 \*\*   
## Student 0.1138478 0.0655441 1.737 0.08272 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6888 on 936 degrees of freedom  
## Multiple R-squared: 0.183, Adjusted R-squared: 0.1768   
## F-statistic: 29.94 on 7 and 936 DF, p-value: < 2.2e-16  
##   
##   
## Response GHQ\_12 :  
##   
## Call:  
## lm(formula = GHQ\_12 ~ Openness + Restraint + Transcendence +   
## Age + Gender + Work + Student, data = DB)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.6148 -2.9404 0.1026 2.8899 13.8171   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 24.748856 1.592292 15.543 < 2e-16 \*\*\*  
## Openness -0.006272 0.014608 -0.429 0.66777   
## Restraint 0.054324 0.021996 2.470 0.01370 \*   
## Transcendence -0.096388 0.011626 -8.290 3.89e-16 \*\*\*  
## Age 0.032541 0.012765 2.549 0.01095 \*   
## Gender -1.068207 0.349031 -3.060 0.00227 \*\*   
## Work 0.183346 0.098921 1.853 0.06413 .   
## Student 1.120192 0.438691 2.553 0.01082 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.61 on 936 degrees of freedom  
## Multiple R-squared: 0.1228, Adjusted R-squared: 0.1162   
## F-statistic: 18.71 on 7 and 936 DF, p-value: < 2.2e-16  
##   
##   
## Response SEC :  
##   
## Call:  
## lm(formula = SEC ~ Openness + Restraint + Transcendence + Age +   
## Gender + Work + Student, data = DB)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.8660 -2.2331 -0.0726 2.2582 13.0752   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.513246 1.192937 1.269 0.2049   
## Openness 0.011424 0.010944 1.044 0.2968   
## Restraint -0.007174 0.016479 -0.435 0.6634   
## Transcendence 0.103471 0.008710 11.879 < 2e-16 \*\*\*  
## Age 0.006810 0.009563 0.712 0.4766   
## Gender 1.044523 0.261493 3.994 6.99e-05 \*\*\*  
## Work -0.152525 0.074111 -2.058 0.0399 \*   
## Student -0.408868 0.328665 -1.244 0.2138   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.454 on 936 degrees of freedom  
## Multiple R-squared: 0.2551, Adjusted R-squared: 0.2496   
## F-statistic: 45.8 on 7 and 936 DF, p-value: < 2.2e-16

## Conclusion

In conclusion , the following models explain the most variations in our response ,given the data at hand .

1. SEC ~ 1.513246 + Openness + Restraint +Transcendence + Age + Gender + Work + Student
2. GHQ\_12 ~ 24.748856 + Openness + Restraint + Transcendence + Age + Gender + Work + Student
3. DASS\_21 ~ 4.0519931+Openness + Restraint + Transcendence + Age + Gender + Work + Student

These models explain 25% , 11.6% ,and 17% of the variations ,respectively .