#factor analysis WVS all

###FA all

#2019

getwd()

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2019")

library(readxl)

wvs2019 <- read\_excel("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2019/testdata2.xlsx")

wvs2019 = as.data.frame(wvs2019)

library(psych)

library(tidyverse)

#https://www.datanovia.com/en/lessons/select-data-frame-columns-in-r/

wvs2019\_2 <- wvs2019 %>%select(Q106 ,

Q107 ,

Q108 ,

Q109 ,

Q110 ,

Q112 ,

Q120 ,

Q143 ,

Q158 ,

Q159 ,

Q160 ,

Q161 ,

Q162 ,

Q177 ,

Q178 ,

Q179 ,

Q180 ,

Q181 ,

Q182 ,

Q183 ,

Q184 ,

Q185 ,

Q186 ,

Q187 ,

Q188 ,

Q189 ,

Q190 ,

Q191 ,

Q192 ,

Q193 ,

Q194 ,

Q195 ,

Q234A ,

Q240 ,

Q241 ,

Q242 ,

Q243 ,

Q244 ,

Q245 ,

Q246 ,

Q247 ,

Q248 ,

Q249 ,

Q250 ,

Q251 ,

Q252 ,

Q286 ,

Q287 ,

Q288 ,

Q288R ,

Q291G2 ,

Q291G3 ,

Q291G4 ,

Q291G5 ,

Q291G6 ,

Q291P1 ,

Q291P2 ,

Q291P3 ,

Q291P4 ,

Q291P5 ,

Q291P6 ,

Q292A ,

Q292B ,

Q292C ,

Q292D ,

Q292E ,

Q292F ,

Q292G ,

Q292H ,

Q292I ,

Q292J ,

Q292K ,

Q292L ,

Q292M ,

Q292N ,

Q292O ,

Q293 ,

Q35 ,

Q36 ,

Q39 ,

Q40 ,

Q41 ,

Q47 ,

Q48 ,

Q49 ,

Q50)

#detach(wvs2019\_2)

attach(wvs2019\_2)

class(wvs2019\_2)

#subset middle class

#d <- as.numeric(c(1:5))

#d <- as.data.frame(d)

#d

#wvs2019\_2m <- subset(wvs2019\_2, Q287>1 & Q287<4)

#wvs2019\_2m = subset(wvs2019\_2, select = -c(Q287, Q288, Q288R))

wvs2019\_2m = subset(wvs2019\_2)

#descriptive

desc <- describe(wvs2019\_2m)

desc2 <- desc%>%select(skew,kurtosis)

desc2

mardia(wvs2019\_2, na.rm=TRUE, plot=TRUE)

out=outlier(wvs2019\_2, bad=5, cex=.5, plot=T, na.rm=TRUE, bg=c("blue"),

pch=21, ylab="D2", ylim=c(0,500))

#pearson

wvs19cor = cor(wvs2019\_2m)

#visualize cor>.3

library(qgraph)

qgraph(wvs19cor,cut=.30,details=TRUE,posCol="darkgreen",negCol="red",

labels=names(wvs19cor))

#correlation plot from the psych package to see corr > .30

corPlot(wvs19cor,diag=F,zlim=c(.3,1),upper=F,numbers=TRUE,cex.axis=.5)

#DETERMINE NUMBER OF CORRELATIONS ABOVE .30

#also chekch for Singularity - too high correlation (r=1).

##create correlation matrix from raw data

wvs19cor = cor(wvs2019\_2m)

##compute number of coef>=.30 off-diagonal

BigR=sum(wvs19cor>=abs(.30) & wvs19cor<abs(1.0),na.rm=T)/2

print(BigR)

#BigR = 41

##Check for multicollinearity

#if determinant of cor matrix is >0.00001 then multicollinearity is probably not a problem

det(cor(wvs2019\_2m))

det(wvs19cor)

#KMO

KMO(wvs2019\_2)

#Bartlett's

cortest.bartlett(wvs2019\_2,n=1200)

#consider:

#https://stackoverflow.com/questions/15215457/standardize-data-columns-in-r

##How many factors to retain?

#PATTERN MATRIX FOR SOLUTION WITH EIGHT

#FACTORS FROM PSYCH PACKAGE

f8=fa(wvs2019\_2,nfactors=8,SMC=TRUE,min.err=0.001,max.iter=1000,fm="ml",rotate="none",n.obs=1200)

f8out <- print(f8,sort=TRUE, cut=.3, digits=2)

#PARALLEL ANALYSIS (PA) WITH PSYCH PACKAGE

#PA with 500 repetitions

#For correlation matrix the n.pbs must be added: n.obs=152

#compare eigen of simulated and actual

pawvs19=fa.parallel(wvs2019\_2m,fa="pc",n.iter = 500,ylab="Eigenvalues",quant=.50)

print(pawvs19)

#13 components

#fa="pc" - extraction method=PCA

#fa="fa" - extraction method=common factor extraction

#quant = comparison standard, here = 50th percentile

#only 1 component (eigenvalue=7.90) is sufficient

#MAP WITH PSYCH PACKAGE

vss(wvs2019\_2m,rotate="none", fm="pc", plot=FALSE, n.obs=1200)

#The lowest MAP value identifies the number of factors to retain. In this

#case, MAP reaches a minimum at two factors

#lowest MAP=?

#indicates 5 to 8 factors

#SCREE PLOT WITH PSYCH PACKAGE

#display scree plot from both reduced and unreduced corr matrices

scree(wvs2019\_2m,pc=TRUE,factors=TRUE,hline="-1",main="Scree Plot")

#how many factors?

#scree = 2-4

#pa = 13

#MAP = 5 - 8

#################

# 6 factor model, promax rotation, ML extraction, save residuals

#missing data can be imputed with mean (impute="mean") or median

#(impute="median"). Default vaues for iteration (min.err, max.iter)

#and initial communality estimate (SMC)

#record output

#sink(file = "Out1.txt", split = TRUE, append = FALSE)

f6=fa(wvs2019\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,cut=.3,sort=TRUE)

#record

sink(file = "f6\_all.txt", split = TRUE, append = FALSE)

print(sort=TRUE,digits=3, cut=.3,f6$Structure)

sink()

#remove loading <.3

wvs2019\_2m = subset(wvs2019\_2m, select = -c(Q177,

Q143,

Q159,

Q158,

Q240,

Q252,

Q48,

Q106,

Q293,

Q120,

Q49,

Q40,

Q107,

Q108,

Q234A,

Q112,

Q39,

Q109,

Q112,

Q39,

Q109,

Q41,

Q292K,

Q36,

Q286,

Q50,

Q291G3,

Q161,

Q110,

Q35,

Q162,

Q47))

f6=fa(wvs2019\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,cut=.3,sort=TRUE)

wvs2019\_2m = subset(wvs2019\_2m, select = -c(Q291P3, Q292H, Q291P2, Q160))

f6=fa(wvs2019\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,cut=.3,sort=TRUE)

wvs2019\_2m = subset(wvs2019\_2m, select = -c(Q291G2))

f6=fa(wvs2019\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,cut=.3,sort=TRUE)

f4=fa(wvs2019\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3,cut=.3,sort=TRUE)

wvs2019\_2m = subset(wvs2019\_2m, select = -c(Q292N,

Q292O,

Q292F,

Q292B,

Q292C,

Q292A,

Q292E,

Q291G5,

Q291P5,

Q292G,

Q292D,

Q292J,

Q292L,

Q292M,

Q292I))

f4=fa(wvs2019\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

sink(file = "wvs2019f4\_all.txt", split = TRUE, append = FALSE)

print(digits=3,cut=.3,sort=TRUE, f4$Structure)

sink()

#ML2 = justifiable

#ML4 = democracy

#ML3 = trust in govt

#ML1 = SES

#get crosstab with Q287

###############################

##############################

############################

#1996

getwd()

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/1996")

library(readxl)

wvs1996 <- read\_excel("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/1996/testdata2.xlsx")

wvs1996 = as.data.frame(wvs1996)

library(psych)

library(tidyverse)

#https://www.datanovia.com/en/lessons/select-data-frame-columns-in-r/

wvs1996\_2 <- wvs1996%>%select(V10 ,

V101 ,

V102 ,

V103 ,

V11 ,

V123 ,

V125 ,

V126 ,

V127 ,

V128 ,

V129 ,

V135 ,

V136 ,

V138 ,

V139 ,

V140 ,

V141 ,

V142 ,

V143 ,

V144 ,

V145 ,

V146 ,

V147 ,

V148 ,

V149\_06 ,

V150 ,

V154 ,

V155 ,

V156 ,

V157 ,

V177 ,

V192 ,

V193 ,

V194 ,

V196 ,

V197 ,

V198 ,

V199 ,

V200 ,

V201 ,

V202 ,

V217 ,

V220 ,

V221 ,

V225 ,

V226 ,

V4 ,

V5 ,

V6 ,

V61 ,

V63 ,

V64 ,

V65 ,

V66 ,

V7 ,

V70 ,

V8 ,

V9 ,

V99)

attach(wvs1996\_2)

class(wvs1996\_2)

wvs1996\_2m = subset(wvs1996\_2)

#descriptive

desc <- describe(wvs1996\_2m)

desc2 <- desc%>%select(skew,kurtosis)

desc2

#V4 is highly skewed

mardia(wvs1996\_2m, na.rm=TRUE, plot=TRUE)

out=outlier(wvs1996\_2m, bad=5, cex=.5, plot=T, na.rm=TRUE, bg=c("blue"),

pch=21, ylab="D2", ylim=c(0,500))

#pearson

wvs96cor = cor(wvs1996\_2m)

#visualize cor>.3

library(qgraph)

qgraph(wvs96cor,cut=.30,details=TRUE,posCol="darkgreen",negCol="red",

labels=names(wvs19cor))

#correlation plot from the psych package to see corr > .30

corPlot(wvs96cor,diag=F,zlim=c(.3,1),upper=F,numbers=TRUE,cex.axis=.5)

#DETERMINE NUMBER OF CORRELATIONS ABOVE .30

#also chekch for Singularity - too high correlation (r=1).

##create correlation matrix from raw data

wvs96cor = cor(wvs1996\_2m)

##compute number of coef>=.30 off-diagonal

BigR=sum(wvs96cor>=abs(.30) & wvs96cor<abs(1.0),na.rm=T)/2

print(BigR)

#BigR = 3

##Check for multicollinearity

#if determinant of cor matrix is >0.00001 then multicollinearity is probably not a problem

det(cor(wvs1996\_2m))

#KMO

KMO(wvs1996\_2)

#Bartlett's

cortest.bartlett(wvs1996\_2,n=1200)

f8=fa(wvs1996\_2m,nfactors=8,SMC=TRUE,min.err=0.001,max.iter=1000,fm="ml",rotate="none",n.obs=568)

f8out <- print(f8,sort=TRUE, digits=2)

#PARALLEL ANALYSIS (PA) WITH PSYCH PACKAGE

#PA with 500 repetitions

#For correlation matrix the n.pbs must be added: n.obs=152

#compare eigen of simulated and actual

pawvs01=fa.parallel(wvs1996\_2m,fa="pc",n.iter = 500,ylab="Eigenvalues",quant=.50)

print(pawvs01)

#suggests 12 factors

#fa="pc" - extraction method=PCA

#fa="fa" - extraction method=common factor extraction

#quant = comparison standard, here = 50th percentile

#only 1 component (eigenvalue=7.90) is sufficient

#MAP WITH PSYCH PACKAGE

vss(wvs1996\_2m,rotate="none", fm="pc", plot=FALSE, n.obs=1200)

#The lowest MAP value identifies the number of factors to retain. In this

#case, MAP reaches a minimum at two factors

#lowest MAP=?

#indicates 3 to 4 factors

#SCREE PLOT WITH PSYCH PACKAGE

#display scree plot from both reduced and unreduced corr matrices

scree(wvs1996\_2m,pc=TRUE,factors=TRUE,hline="-1",main="Scree Plot")

#how many factors?

#scree = 2-3

#pa = 12

#MAP = 3 - 4

#sink(file = "Out1.txt", split = TRUE, append = FALSE)

f6=fa(wvs1996\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,sort=TRUE, cut=.3)

wvs1996\_2m = subset(wvs1996\_2m, select = -c(V135,

V101,

V7,

V123,

V103,

V126,

V9,

V6,

V70,

V226,

V177,

V221,

V157,

V63,

V128,

V4,

V129,

V5,

V11,

V10,

V225,

V125,

V8,

V127,

V99,

V102,

V220,

V61))

f6=fa(wvs1996\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,sort=TRUE, cut=.3)

#indicates overfactoring

f5=fa(wvs1996\_2m,nfactors=5,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f5,digits=3,sort=TRUE, cut=.3)

#indicates overfactoring

f4=fa(wvs1996\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3,sort=TRUE, cut=.3)

wvs1996\_2m = subset(wvs1996\_2m, select = -c(V154, V155))

f4=fa(wvs1996\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3,sort=TRUE, cut=.3)

wvs1996\_2m = subset(wvs1996\_2m, select = -c(V156))

f4=fa(wvs1996\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3,sort=TRUE, cut=.3)

#seems like f3 is reasonable. the deleted vars that follow overfactors with others and belongs to ML4

wvs1996\_2m = subset(wvs1996\_2m, select = -c(V150, V149\_06, V148, V66))

f3=fa(wvs1996\_2m,nfactors=3,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f3,digits=3,sort=TRUE, cut=.3)

wvs1996\_2m = subset(wvs1996\_2m, select = -c(V217))

f3=fa(wvs1996\_2m,nfactors=3,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

sink(file = "wvs1996f3\_all.txt", split = TRUE, append = FALSE)

print(digits=3,sort=TRUE, cut=.3, f3$Structure)

sink()

#ML1 = Justifiable

#ML2 = confidence in organizations/institutions

#ML3 = satisfaction

#better model

#####################

###################

#2001

###Codes for WVS 2001

getwd()

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2001")

library(readxl)

wvs2001 <- read\_excel("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2001/testdata2.xlsx")

wvs2001 = as.data.frame(wvs2001)

library(psych)

library(tidyverse)

#https://www.datanovia.com/en/lessons/select-data-frame-columns-in-r/

wvs2001\_2 <- wvs2001 %>%select(V100 ,

V102 ,

V11 ,

V113 ,

V116 ,

V118 ,

V119 ,

V12 ,

V133 ,

V139 ,

V141 ,

V142 ,

V143 ,

V144 ,

V147 ,

V148 ,

V149 ,

V150 ,

V151 ,

V152 ,

V153 ,

V154 ,

V155 ,

V156 ,

V157 ,

V158 ,

V159 ,

V160C ,

V162 ,

V164 ,

V165 ,

V166 ,

V167 ,

V173 ,

V204 ,

V205 ,

V206 ,

V207 ,

V208 ,

V209 ,

V210 ,

V211 ,

V212 ,

V213 ,

V229 ,

V234 ,

V235 ,

V236 ,

V4 ,

V5 ,

V6 ,

V7 ,

V78 ,

V79 ,

V8 ,

V80 ,

V81 ,

V82 ,

V9 ,

V99)

#detach(wvs2001\_2)

attach(wvs2001\_2)

class(wvs2001\_2)

#subset middle class

#d <- as.numeric(c(1:5))

#d <- as.data.frame(d)

#d

#wvs2001\_2m <- subset(wvs2001\_2, V235>1 & V235<4)

#wvs2001\_2m = subset(wvs2001\_2m, select = -c(V235))

wvs2001\_2m = subset(wvs2001\_2)

#descriptive

desc <- describe(wvs2012\_2m)

desc2 <- desc%>%select(skew,kurtosis)

desc2

#V4 is highly skewed

mardia(wvs2001\_2m, na.rm=TRUE, plot=TRUE)

out=outlier(wvs2001\_2m, bad=5, cex=.5, plot=T, na.rm=TRUE, bg=c("blue"),

pch=21, ylab="D2", ylim=c(0,500))

#pearson

wvs01cor = cor(wvs2001\_2m)

#visualize cor>.3

library(qgraph)

qgraph(wvs01cor,cut=.30,details=TRUE,posCol="darkgreen",negCol="red",

labels=names(wvs01cor))

#correlation plot from the psych package to see corr > .30

corPlot(wvs01cor,diag=F,zlim=c(.3,1),upper=F,numbers=TRUE,cex.axis=.5)

#DETERMINE NUMBER OF CORRELATIONS ABOVE .30

#also chekch for Singularity - too high correlation (r=1).

##create correlation matrix from raw data

wvs01cor = cor(wvs2001\_2m)

##compute number of coef>=.30 off-diagonal

BigR=sum(wvs01cor>=abs(.30) & wvs01cor<abs(1.0),na.rm=T)/2

print(BigR)

#BigR = 79

##Check for multicollinearity

#if determinant of cor matrix is >0.00001 then multicollinearity is probably not a problem

det(cor(wvs2001\_2m))

#KMO

KMO(wvs2001\_2)

#Bartlett's

cortest.bartlett(wvs2001\_2,n=1200)

#consider:

#https://stackoverflow.com/questions/15215457/standardize-data-columns-in-r

##How many factors to retain?

#PATTERN MATRIX FOR SOLUTION WITH EIGHT

#FACTORS FROM PSYCH PACKAGE

f8=fa(wvs2001\_2m,nfactors=8,SMC=TRUE,min.err=0.001,max.iter=1000,fm="ml",rotate="none",n.obs=1200)

f8out <- print(f8,sort=TRUE, cut=.3,digits=2)

#PARALLEL ANALYSIS (PA) WITH PSYCH PACKAGE

#PA with 500 repetitions

#For correlation matrix the n.pbs must be added: n.obs=152

#compare eigen of simulated and actual

pawvs01=fa.parallel(wvs2001\_2m,fa="pc",n.iter = 500,ylab="Eigenvalues",quant=.50)

print(pawvs01)

#suggests 14 factors

#fa="pc" - extraction method=PCA

#fa="fa" - extraction method=common factor extraction

#quant = comparison standard, here = 50th percentile

#only 1 component (eigenvalue=7.90) is sufficient

#MAP WITH PSYCH PACKAGE

vss(wvs2001\_2m,rotate="none", fm="pc", plot=FALSE, n.obs=1200)

#The lowest MAP value identifies the number of factors to retain. In this

#case, MAP reaches a minimum at two factors

#lowest MAP=?

#indicates 5 factors

#SCREE PLOT WITH PSYCH PACKAGE

#display scree plot from both reduced and unreduced corr matrices

scree(wvs2001\_2m,pc=TRUE,factors=TRUE,hline="-1",main="Scree Plot")

###f6

f6=fa(wvs2001\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,sort=TRUE, cut=.3)

wvs2001\_2m = subset(wvs2001\_2m, select = -c(V147,

V7,

V133,

V9,

V99,

V229,

V139,

V12,

V82,

V6,

V141,

V5,

V143,

V79,

V102,

V100,

V142,

V118,

V78,

V113,

V116,

V144,

V173,

V4,

V167,

V8))

f6=fa(wvs2001\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,sort=TRUE, cut=.3)

wvs2001\_2m = subset(wvs2001\_2m, select =-c(V119))

f6=fa(wvs2001\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,sort=TRUE, cut=.3)

wvs2001\_2m = subset(wvs2001\_2m, select =-c(V159))

f6=fa(wvs2001\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3,sort=TRUE, cut=.3)

#ML2 = confidence in institutions

#ML1 = Justifiable

#ML5 = Justifiable

#ML3 = Satisfaction and SES

#ML4 = confidence in institutions (asean and UN)

#ML6 = governance

#ML2 & ML4; ML1 and ML5 <- signs of overfactoring; try f4

f4=fa(wvs2001\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3,sort=TRUE, cut=.3)

wvs2001\_2m = subset(wvs2001\_2m, select =-c(V235))

f4=fa(wvs2001\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3,sort=TRUE, cut=.3)

wvs2001\_2m = subset(wvs2001\_2m, select =-c(V166, V164, V165, V234))

f4=fa(wvs2001\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3,sort=TRUE, cut=.3)

#better model--no overfactoring, slightly lower cum var

#ML2 = Confidence in institutions

#ML1 = Justifiable

#ML4 = Justifiable (has to do with money)

#ML3 = Satisfaction and SES

#try f3

f3=fa(wvs2001\_2m,nfactors=3,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

sink(file = "wvs2001f3\_all.txt", split = TRUE, append = FALSE)

print(digits=3,sort=TRUE, cut=.3, f3$Structure)

sink()

#ML2 = Confidence in institutions

#ML1 = Justifiable

#ML3 = Satisfaction and SES

#not bad. justifiable in 1 factor, lower cum var

#only year so far where SES loaded significantly; but still not "V235"

###Codes for WVS 2012

getwd()

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2012")

library(readxl)

wvs2012 <- read\_excel("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2012/testdata2.xlsx")

wvs2012 = as.data.frame(wvs2012)

library(psych)

library(tidyverse)

#https://www.datanovia.com/en/lessons/select-data-frame-columns-in-r/

wvs2012\_2 <- wvs2012 %>%select(V10 ,

V100 ,

V108 ,

V109 ,

V11 ,

V110 ,

V111 ,

V112 ,

V112 ,

V114 ,

V115 ,

V116 ,

V117 ,

V118 ,

V118 ,

V119 ,

V120 ,

V121 ,

V122 ,

V123 ,

V124 ,

V125\_09 ,

V126 ,

V127 ,

V128 ,

V129 ,

V130 ,

V131 ,

V132 ,

V133 ,

V134 ,

V135 ,

V136 ,

V137 ,

V138 ,

V139 ,

V140 ,

V141 ,

V142 ,

V170 ,

V171 ,

V172 ,

V173 ,

V174 ,

V175 ,

V181 ,

V182 ,

V189 ,

V190 ,

V191 ,

V192 ,

V193 ,

V194 ,

V195 ,

V196 ,

V198 ,

V199 ,

V200 ,

V201 ,

V202 ,

V203 ,

V203A ,

V204 ,

V205 ,

V206 ,

V207 ,

V207A ,

V208 ,

V209 ,

V210 ,

V228J ,

V229 ,

V23 ,

V230 ,

V238,

V248 ,

V4 ,

V45 ,

V46 ,

V47 ,

V49 ,

V5 ,

V50 ,

V51 ,

V52 ,

V53 ,

V54 ,

V55 ,

V59 ,

V6 ,

V7 ,

V8 ,

V84 ,

V9 ,

V95 ,

V96 ,

V97 ,

V98 ,

V99)

#detach(wvs2012\_2)

attach(wvs2012\_2)

class(wvs2012\_2)

#subset middle class

#d <- as.numeric(c(1:5))

#d <- as.data.frame(d)

#d

#wvs2012\_2m <- subset(wvs2012\_2, V238>1 & V238<4)

#wvs2012\_2m = subset(wvs2012\_2m, select = -c(V238))

wvs2012\_2m = subset(wvs2012\_2)

#descriptive

desc <- describe(wvs2012\_2m)

desc2 <- desc%>%select(skew,kurtosis)

desc2

#V4 is highly skewed

mardia(wvs2019\_2m, na.rm=TRUE, plot=TRUE)

out=outlier(wvs2019\_2m, bad=5, cex=.5, plot=T, na.rm=TRUE, bg=c("blue"),

pch=21, ylab="D2", ylim=c(0,500))

#pearson

wvs12cor = cor(wvs2012\_2m)

#visualize cor>.3

library(qgraph)

qgraph(wvs12cor,cut=.30,details=TRUE,posCol="darkgreen",negCol="red",

labels=names(wvs12cor))

#correlation plot from the psych package to see corr > .30

corPlot(wvs12cor,diag=F,zlim=c(.3,1),upper=F,numbers=TRUE,cex.axis=.5)

#DETERMINE NUMBER OF CORRELATIONS ABOVE .30

#also chekch for Singularity - too high correlation (r=1).

##create correlation matrix from raw data

wvs12cor = cor(wvs2012\_2m)

##compute number of coef>=.30 off-diagonal

BigR=sum(wvs12cor>=abs(.30) & wvs12cor<abs(1.0),na.rm=T)/2

print(BigR)

#BigR = 150

##Check for multicollinearity

#if determinant of cor matrix is >0.00001 then multicollinearity is probably not a problem

det(cor(wvs2012\_2m))

#KMO

KMO(wvs2012\_2)

#Bartlett's

cortest.bartlett(wvs2012\_2,n=1200)

#consider:

#https://stackoverflow.com/questions/15215457/standardize-data-columns-in-r

##How many factors to retain?

#PATTERN MATRIX FOR SOLUTION WITH EIGHT

#FACTORS FROM PSYCH PACKAGE

f8=fa(wvs2012\_2m,nfactors=8,SMC=TRUE,min.err=0.001,max.iter=1000,fm="ml",rotate="none",n.obs=733)

f8out <- print(f8,sort=TRUE, cut = .3, digits=2)

#PARALLEL ANALYSIS (PA) WITH PSYCH PACKAGE

#PA with 500 repetitions

#For correlation matrix the n.pbs must be added: n.obs=152

#compare eigen of simulated and actual

pawvs12=fa.parallel(wvs2012\_2m,fa="pc",n.iter = 500,ylab="Eigenvalues",quant=.50)

print(pawvs12)

#suggests 18 factors

#fa="pc" - extraction method=PCA

#fa="fa" - extraction method=common factor extraction

#quant = comparison standard, here = 50th percentile

#only 1 component (eigenvalue=7.90) is sufficient

#MAP WITH PSYCH PACKAGE

vss(wvs2012\_2m,rotate="none", fm="pc", plot=FALSE, n.obs=1200)

#The lowest MAP value identifies the number of factors to retain. In this

#case, MAP reaches a minimum at two factors

#lowest MAP=?

#indicates 8 factors

#SCREE PLOT WITH PSYCH PACKAGE

#display scree plot from both reduced and unreduced corr matrices

scree(wvs2012\_2m,pc=TRUE,factors=TRUE,hline="-1",main="Scree Plot")

#indicates 2-3 factors

f6=fa(wvs2012\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3, cut = .3, sort=TRUE)

wvs2012\_2m = subset(wvs2012\_2m, select = -c(V198,

V108,

V130,

V84,

V129,

V127,

V142,

V9,

V8,

V132,

V193,

V192,

V135,

V55,

V195,

V7,

V196,

V100,

V97,

V98,

V46,

V194,

V128,

V99,

V228J,

V96,

V59,

V23,

V10,

V11,

V5,

V4,

V45,

V238,

V6,

V230,

V181,

V49,

V229))

f6=fa(wvs2012\_2m,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f6,digits=3, cut = .3, sort=TRUE)

#there is indication of overfactoring

#try f4

f4=fa(wvs2012\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3, cut = .3, sort=TRUE)

wvs2012\_2m = subset(wvs2012\_2m, select = -c(V182, V54, V47,

V95, V171, V173,

V175, V189, V174,

V170, V172, V52,

V53, V190, V51,

V191, V50))

f4=fa(wvs2012\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3, cut = .3, sort=TRUE)

wvs2012\_2m = subset(wvs2012\_2m, select = -c(V141))

f4=fa(wvs2012\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3, cut = .3, sort=TRUE)

wvs2012\_2m = subset(wvs2012\_2m, select = -c(V248))

f4=fa(wvs2012\_2m,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f4,digits=3, cut = .3, sort=TRUE)

#ML1 = justifiable

#ML2 = confidence in institutions

#ML3 = Democracy

#ML4 = confidence in institutions (Dev't orgs)

#try f3

f3=fa(wvs2012\_2m,nfactors=3,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=1200)

print(f3,digits=3, cut = .3, sort=TRUE)

sink(file = "wvs2012all\_f3.txt", split = TRUE, append = FALSE)

print(sort=TRUE,digits=3, cut=.3,f3$Structure)

sink()

#ML1 = justifiable

#ML2 = confidence in institutions

#ML3 = Democracy

#lower cum var than f4

####################

###################

##################

#Ambisyon natin codes

#library(foreign)

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/amb2040")

library(psych)

library(haven)

library(tidyverse)

amb2040 <- read\_sav("amb2040.sav")

vars <- ls(amb2040)

sink(file = "ambvars.txt", split = TRUE, append = FALSE)

options(max.print=1600)

print(vars)

sink()

describe(amb2040[,1:5])

amb2040\_2 <- amb2040%>%select(vq10, vq40, vq6, vq9,

vq7, vq26, vq48.1, vq48.2,

vq48.3, vq48.4, vq48.5, vq48.6,

vq29, vq29, vq38, q42coded,

vq1, f5coded, vq63, vq13,

vq35, vq36, vq41, vq39,

vq17a, vq17b, vq17c, vq8,

vq54, vq55, vq45, eco2)

#desc <- describe(amb2040\_2)

#desc

#colnames(amb2040\_2)

#table(amb2040$eco2)

#f5coded and vq55 = high skew, kurtosis

#weights not needed?

amb2040\_2 = subset(amb2040\_2, select = -c(f5coded, vq55))

#amb2040\_2 <- subset(amb2040\_2, eco2>1 & eco2<4)

#amb2040\_2 = subset(amb2040\_2, select = -c(eco2))

amb2040\_2cor = cor(amb2040\_2)

BigR=sum(amb2040\_2cor>=abs(.30) & amb2040\_2cor<abs(1.0),na.rm=T)/2

print(BigR)

#41

det(cor(amb2040\_2))

KMO(amb2040\_2)

cortest.bartlett(amb2040\_2,n=10000)

#PARALLEL ANALYSIS (PA) WITH PSYCH PACKAGE

#PA with 500 repetitions

#For correlation matrix the n.pbs must be added: n.obs=152

#compare eigen of simulated and actual

paamb=fa.parallel(amb2040\_2,fa="pc",n.iter = 500,ylab="Eigenvalues",quant=.50)

print(paamb)

#pa = 6 factors

#MAP WITH PSYCH PACKAGE

vss(amb2040\_2,rotate="none", fm="pc", plot=FALSE, n.obs=10000)

#suggests 4 factors

scree(amb2040\_2,pc=TRUE,factors=TRUE,hline="-1",main="Scree Plot")

#suggests 2-3 factors

#f6

f6=fa(amb2040\_2,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=10000)

print(f6,digits=3,sort=TRUE, cut=.3)

amb2040\_2 = subset(amb2040\_2, select = -c(vq8, vq63, vq1, vq54, vq45))

f6=fa(amb2040\_2,nfactors=6,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=10000)

sink(file = "f6amb\_all.txt", split = TRUE, append = FALSE)

print(digits=3,sort=TRUE, cut=.3, f6$Structure)

sink()

#try f5

f4=fa(amb2040\_2,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=10000)

print(digits=3,sort=TRUE, cut=.3, f4$Structure)

#seems there is overlap of loadings but they make sense.

#confidence and income level have different signs (+ and -, respectively)

#remove loading <.3

amb2040\_2 = subset(amb2040\_2, select = -c(vq17a, vq17b, vq17c, vq38, q42coded, vq41))

f4=fa(amb2040\_2,nfactors=4,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=10000)

print(digits=3,sort=TRUE, cut=.3, f4$Structure)

#high cum var

#large overlap

#try f3

#large overlap

#try f2, varimax

f2=fa(amb2040\_2,nfactors=2,rotate="varimax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=10000)

print(digits=3,sort=TRUE, cut=.3, f2$Structure)

#remove loading <.3

amb2040\_2 = subset(amb2040\_2, select = -c(vq40, vq48.1, vq13, vq35, vq36, vq39))

f2=fa(amb2040\_2,nfactors=2,rotate="varimax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=10000)

print(digits=3,sort=TRUE, cut=.3, f2$Structure)

amb2040\_2 = subset(amb2040\_2, select = -c(eco2))

f2=fa(amb2040\_2,nfactors=2,rotate="varimax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=10000)

sink(file = "ambf2allvarimax.txt", split = TRUE, append = FALSE)

print(digits=3,sort=TRUE, cut=.3, f2$Structure)

sink()

#attribute

#1996

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/1996")

library(readxl)

wvs1996 <- read\_excel("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/1996/testdata2.xlsx")

wvs1996 = as.data.frame(wvs1996)

#2001

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2001")

library(readxl)

wvs2001 <- read\_excel("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2001/testdata2.xlsx")

wvs2001 = as.data.frame(wvs2001)

#2012

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2012")

library(readxl)

wvs2012 <- read\_excel("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2012/testdata2.xlsx")

wvs2012 = as.data.frame(wvs2012)

#2019

setwd("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2019")

library(readxl)

wvs2019 <- read\_excel("D:/Users/Erwin/OneDrive - University of the Philippines/CSWCD/SD/SD 400/Dataset/WVS/2019/testdata2.xlsx")

wvs2019 = as.data.frame(wvs2019)

#1996

library(gmodels)

wvs1996$V226[wvs1996$V226 == -1] <- NA

CrossTable(wvs1996$V217R, wvs1996$V226, chisq = TRUE) #\*\*educ recoded

CrossTable(wvs1996$V217, wvs1996$V217R, chisq = TRUE) #\*\*educ

CrossTable(wvs1996$V217, wvs1996$V226, chisq = TRUE) #\*\*educ

CrossTable(wvs1996$V220, wvs1996$V226, chisq = TRUE) #\*\*employment status

CrossTable(wvs1996$V221, wvs1996$V226, chisq = TRUE) #\*\*profession

CrossTable(wvs1996$V225, wvs1996$V226, chisq = TRUE) #\*\*savings

wvs2001$V235[wvs2001$V235 == -1] <- NA

wvs2001$V226R[wvs2001$V226R == -1] <- NA

CrossTable(wvs2001$V226R, wvs2001$V235, chisq = TRUE) #\*\*educ recoded

CrossTable(wvs2001$V226, wvs2001$V226R, chisq = TRUE) #\*\*educ

CrossTable(wvs2001$V226, wvs2001$V235, chisq = TRUE) #\*\*educ

CrossTable(wvs2001$V229, wvs2001$V235, chisq = TRUE) #\*\*employment status

CrossTable(wvs2001$V230, wvs2001$V235, chisq = TRUE) #\*\*profession

CrossTable(wvs2001$V233, wvs2001$V235, chisq = TRUE) #\*\*profession

CrossTable(wvs2001$V234, wvs2001$V235, chisq = TRUE) #\*\*savings

wvs2012$V238[wvs2012$V238 == -1] <- NA

wvs2012$V248R[wvs2012$V248 == -1] <- NA

CrossTable(wvs2012$V248R, wvs2012$V238, chisq = TRUE) #\*\*educ recoded

CrossTable(wvs2012$V248, wvs2012$V248R, chisq = TRUE) #\*\*educ

CrossTable(wvs2012$V248, wvs2012$V238, chisq = TRUE) #\*\*educ

CrossTable(wvs2012$V229, wvs2012$V238, chisq = TRUE) #\*\*employment status

CrossTable(wvs2012$V230, wvs2012$V238, chisq = TRUE) #\*\*profession - not available. changed to gov't, private, NGO (V230)

CrossTable(wvs2012$V237, wvs2012$V238, chisq = TRUE) #\*\*savings

wvs2019$Q287[wvs2019$Q287 == -1] <- NA

wvs2019$Q248R[wvs2019$Q248 == -1] <- NA

CrossTable(wvs2019$Q275R, wvs2019$Q287, chisq = TRUE) #\*\*educ recoded

CrossTable(wvs2019$Q275, wvs2019$Q275R, chisq = TRUE) #\*\*educ

CrossTable(wvs2019$Q275, wvs2019$Q287, chisq = TRUE) #\*\*educ

CrossTable(wvs2019$Q279, wvs2019$Q287, chisq = TRUE) #\*\*employment status

CrossTable(wvs2019$Q281, wvs2019$Q287, chisq = TRUE) #\*\*profession

CrossTable(wvs2019$Q282, wvs2019$Q287, chisq = TRUE) #\*\*profession

CrossTable(wvs2019$Q286, wvs2019$Q287, chisq = TRUE) #\*\*savings