#analysis of AMB 2040

#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)

#48

det(cor(amb2040\_2))

KMO(amb2040\_2)

cortest.bartlett(amb2040\_2,n=2220)

#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=2220)

#suggests 4 factors

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

#suggests 3-4 factors

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

missing=FALSE,fm="ml",n.obs=2220)

print(f6,digits=3,sort=TRUE, cut=.399)

#record

sink(file = "Outf6.txt", split = TRUE, append = FALSE)

print(sort=TRUE,digits=3, cut=.399,f6$Structure)

sink()

#5-factor model

amb2040\_2 = subset(amb2040\_2, select = -c(vq1, vq63, vq8, vq54, vq45, vq17a, vq17b, vq17c))

f5=fa(amb2040\_2,nfactors=5,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=2220)

print(f5,digits=3,sort=TRUE, cut=.399)

amb2040\_2 = subset(amb2040\_2, select = -c(vq54, vq17b, vq8, vq63, vq1,vq17c, vq17a, vq45))

sink(file = "Outf5.txt", split = TRUE, append = FALSE)

print(sort=TRUE,digits=3, cut=.399,f5$Structure)

sink()

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

missing=FALSE,fm="ml",n.obs=2220)

print(f4,digits=3,sort=TRUE, cut=.399)

#f5-f4 are interesting.

#highest cum var

#but with overfactoring - items 48.6, 48.1 loads in 2 factors.

#consider deleting vars loading<.399 then run 3-factor

amb2040\_2 = subset(amb2040\_2, select = -c(vq36, vq35))

f3=fa(amb2040\_2,nfactors=3,rotate="promax",residuals=TRUE,SMC=TRUE,

missing=FALSE,fm="ml",n.obs=2220)

print(f3,digits=3,sort=TRUE, cut=.399)

sink(file = "Outf3.txt", split = TRUE, append = FALSE)

print(sort=TRUE,digits=3, cut=0,f3$Structure)

sink()

table(amb2040\_2$q42coded)

amb2040\_2 = subset(amb2040\_2, select = -c(q42coded, vq38, vq41, vq39, vq40))

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

missing=FALSE,fm="ml",n.obs=2220)

print(f2,digits=3,sort=TRUE, cut=.399)

sink(file = "Outf2.txt", split = TRUE, append = FALSE)

print(sort=TRUE,digits=3, cut=.399,f2$Structure)

sink()

#f2 seems best

# highest cum var at .61

#factor 1 is educ

#factor 2 is confidence and medical expenses

library(foreign)

library(nnet)

library(ggplot2)

library(reshape2)

amb2040\_2 <- as.data.frame(amb2040\_2)

testml <- multinom(amb2040\_2$eco2 ~ amb2040\_2$vq10 + amb2040\_2$vq40 + amb2040\_2$vq6

+ amb2040\_2$vq9 + amb2040\_2$vq7 + amb2040\_2$vq26 +amb2040\_2$vq48.1

+ amb2040\_2$vq48.2 + amb2040\_2$vq48.3 + amb2040\_2$vq48.4 +amb2040\_2$vq48.5

+ amb2040\_2$vq48.6 + amb2040\_2$vq29 + amb2040\_2$vq38, maxit=1000, data=amb2040\_2, Hess = TRUE)

sum\_ml <- summary(testml)

sum\_ml

###crosstable

#get eco2 again

library(gmodels)

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

CrossTable(amb2040\_2$vq10,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq6,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq9,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq7,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq26,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq29,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq48.3,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq48.4,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq48.5,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq48.6,amb2040\_2$eco2, chisq = TRUE) #

CrossTable(amb2040\_2$vq48.2,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$vq48.1,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$vq35,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$vq36,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$vq40,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$vq38,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$q42coded,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$vq13,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$vq41,amb2040\_2$eco2, chisq = TRUE) #

#CrossTable(amb2040\_2$vq39,amb2040\_2$eco2, chisq = TRUE) #

sink()