CodeReport ST841 TAA Project

Andrew Hollis, Xinyu Zhang, and Qiang Heng

10/26/2020

Introduction

This document is meant to explain and demonstrate the implementation of all analysis performed on the TAA dataset.

Data Transformations

We took the original data and performed several important transformations.

First we read in the initial data file:

```
init_TAA_data<-read.delim("initial_data.txt")
init_TAA_data$Q26_num<-as.numeric(init_TAA_data$Q26_num) #Insure Q26_num is actually numeric
init_TAA_data$Q29_num<-as.numeric(as.character(init_TAA_data$Q29_num)) #Insure Q29_num is actually nume
```

```
## Warning: Ç¿ÖÆ¸Ä±ä¹ý³ÌÖÐ²úÉúÁËNA
```

We then remove the variables that are not as interesting for analysis these include the data from Q42, Q43, Q44, and Q45 which are demographic data about the respondent. We will also compute new summary score for question 6 from the survey, so we remove this as well.

```
removal_list<-c("Q42","Q43","Q44","Q45","Q6_score")
init_TAA_data<-init_TAA_data[, !(names(init_TAA_data) %in%removal_list)]
```

Dealing with Likert Scale Items (Drew)

Questions 5, 10, 23, 30, 31, and 32 are composed of several 5-item Likert scale sub-questions. Each of these questions represents a single underlying conceptual variable. For instance, the 6 subquestions from question 5 together measure the extent to which a business considered various business objectives as they intially considered whether they would try TAA tools. We would like to create a one-dimensional measure that represents the level of this initial TAA consideration for a particular company.

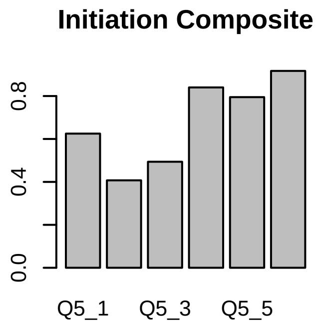
Nonlinear principal components analysis is a methodology that allows us to transform several Likert scale items into the one-dimensional summary metric that captures the maximal amount of variability from the original data. (Linting et al. 2007)

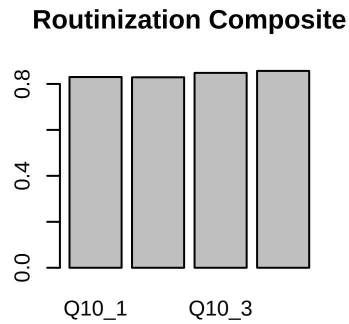
Below, we use R to perform the nonlinear principal components analysis:

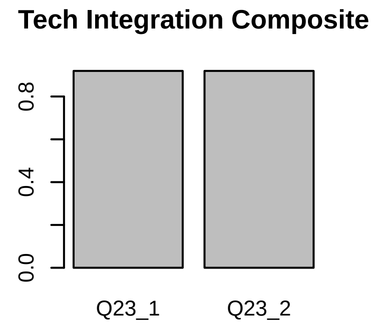
```
library(Gifi)
##Q5
#Set Aside Data for initiation
init_data_names<-c("Q5_1","Q5_2","Q5_3","Q5_4","Q5_5","Q5_6")
init_data<-init_TAA_data[init_data_names]

#Code Missing Values
init_data<-apply(init_data,2,function(x){ifelse(x==6,NA,x)})
```

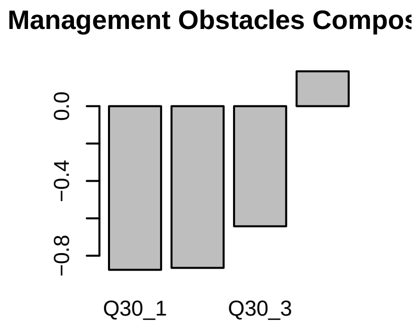
```
#Impute, transform, do PCA
pca_init<-princals(init_data,ndim=1,missing = "a",degrees=2)
init_score<-pca_init$objectscores[,1]

##Q10
#Set Aside Data for routinization
rout_data_names<-c("Q10_1","Q10_2","Q10_3","Q10_4")
rout_data<-init_TAA_data[rout_data_names]

#Code Missing Values
rout_data<-apply(rout_data,2,function(x){ifelse(x==6,NA,x)})

#Impute, transform, do PCA
pca_rout<-princals(rout_data,ndim=1,missing = "a",degrees=2)
rout_score<-pca_rout$objectscores[,1]

##Q23
#Set Aside Data for routinization
integ_data_names<-c("Q23_1","Q23_2")
integ_data<-init_TAA_data[integ_data_names]

#Code Missing Values
integ_data<-apply(integ_data,2,function(x){ifelse(x==6,NA,x)})

#Impute, transform, do PCA
pca_integ<-princals(integ_data,ndim=1,missing = "a",degrees=2)
integ_score<-pca_integ$objectscores[,1]

##Q30
#Set Aside Data for management obstacles
manag_data_names<-c("Q30_1","Q30_2","Q30_3","Q30_4")
manag_data<-init_TAA_data[manag_data_names]

#Code Missing Values
manag_data<-apply(manag_data,2,function(x){ifelse(x==6,NA,x)})

#Impute, transform, do PCA
pca_manag<-princals(manag_data,ndim=1,missing = "a",degrees=2)
manag_score<-pca_manag$objectscores[,1]

##Q31
#Set Aside Data for management obstacles
comp_data_names<-c("Q31_1","Q31_2","Q31_3")
comp_data<-init_TAA_data[comp_data_names]

#Code Missing Values
comp_data<-apply(comp_data,2,function(x){ifelse(x==6,NA,x)})

#Impute, transform, do PCA
pca_comp<-princals(comp_data,ndim=1,missing = "a",degrees=2)
comp_score<-pca_comp$objectscores[,1]

##Q32
```

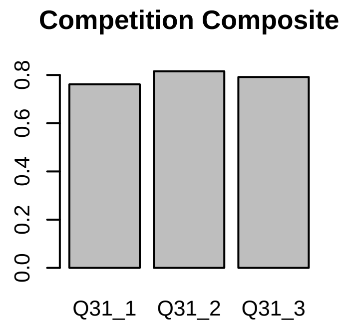
```r
#Set Aside Data for Government Regulation
gov_data_names<-c("Q32_1","Q32_2","Q32_3","Q32_4")
gov_data<-init_TAA_data[gov_data_names]

#Code Missing Values
gov_data<-apply(gov_data,2,function(x){ifelse(x==6,NA,x)})

#Impute, transform, do PCA
pca_gov<-princals(gov_data,ndim=1,missing = "a",degrees=2)
gov_score<-pca_gov$objectscores[,1]

#Add Nonlinear PCA Scores to data
init_TAA_data$init_composite<-init_score
init_TAA_data$rout_composite<-rout_score
init_TAA_data$integrate_composite<-integ_score
init_TAA_data$manag_composite<-manag_score
init_TAA_data$comp_composite<-comp_score
init_TAA_data$gov_composite<-gov_score

#Create Loadings Plots
par(mfrow=c(2,3))
barplot(pca_init$loadings[,1],main="Initiation Composite")
barplot(pca_rout$loadings[,1],main="Routinization Composite")
barplot(pca_integ$loadings[,1],main="Tech Integration Composite")
barplot(pca_manag$loadings[,1],main="Management Obstacles Composite")
barplot(pca_comp$loadings[,1],main="Competition Composite")
barplot(pca_gov$loadings[,1],main="Regulatory Environment Composite")
```

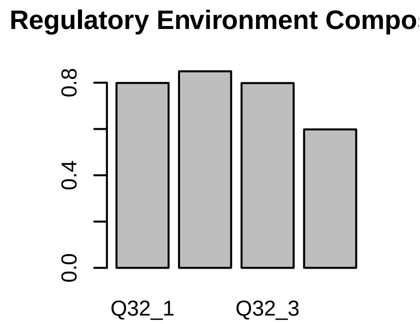













```
#Remove Original Likert Scale Data
removal_list<-c("Q5_1","Q5_2","Q5_3","Q5_4","Q5_5","Q5_6","Q10_1","Q10_2","Q10_3","Q10_4","Q23_1","Q23_
init_TAA_data<-init_TAA_data[, !(names(init_TAA_data) %in%removal_list)]
```

Combining Weighted Variables (Drew)

Q6 and Q22 have survey respondents check several boxes to indicate the level of TAA technology they have adopted and the specific types of TAA they adopted. For Q6, if respondents check the first box, they have adopted emerging TAA technology, if they check the second box they have adopted intermediate TAA technology, and if they check the third box they have adopted advanced TAA technology. We assign a weight of 1 to checkbox 1, 2 to checkbox 2, and 3 to checkbox 3. We create the single adoption score for each company by adding the weights and dividing by the maximum possible cumulative weight, 6.

Similarly, for Q22, if a weight of 0 is assigned to the first box, 1 to the second box, 2 for the 3rd, 4th, 5th, and 10th boxes, 3 for the 6th box, 4 for the 7th box, and 5 for the 8th and 9th boxes. As with the Q6 response, we form a single score by adding the weights and dividing by the maximum potential score, 26.

We standardize both new scores to have zero meana and standard deviation of 1.

We carry out this weighting in R below:

```
#Transform 4's
init_TAA_data[2,9]=3
init_TAA_data[16,7]=1

#Form Adoption Score
adopt_data<-init_TAA_data[c("Q6box_1","Q6box_2","Q6box_3")]
adopt_comp<-apply(adopt_data,1,sum,na.rm=TRUE)/6
```

```
init_TAA_data$adoption_score<-adopt_comp
init_TAA_data$adoption_score<-scale(init_TAA_data$adoption_score)

#Technology Readiness/Capability Score
TAA_capability_data<-init_TAA_data[c("Q22_box1","Q22_box2","Q22_box3","Q22_box4","Q22_box5","Q22_box6",

TAA_capability_score<-rep(0,nrow(TAA_capability_data))
max_score=1+2*4+3+4+5*2
for(i in 1:length(TAA_capability_score)){
  score=0
  if(!is.na(TAA_capability_data[i,1])){
    score=score+0
  }
  if(!is.na(TAA_capability_data[i,2])){
    score=score+1
  }
  if(!is.na(TAA_capability_data[i,3])){
    score=score+2
  }
  if(!is.na(TAA_capability_data[i,4])){
    score=score+2
  }
  if(!is.na(TAA_capability_data[i,5])){
    score=score+2
  }
  if(!is.na(TAA_capability_data[i,6])){
    score=score+3
  }
  if(!is.na(TAA_capability_data[i,7])){
    score=score+4
  }
  if(!is.na(TAA_capability_data[i,8])){
    score=score+5
  }
  if(!is.na(TAA_capability_data[i,9])){
    score=score+5
  }
  if(!is.na(TAA_capability_data[i,10])){
    score=score+2
  }
  TAA_capability_score[i]=score/max_score
}
init_TAA_data$TAA_capability<-TAA_capability_score
init_TAA_data$TAA_capability<-scale(init_TAA_data$TAA_capability)
```

Creating Binary Variable for Q24

Q24 measures the global extent of a company. We transform this variable into a binary variable that is 0 if the company only has domestic offices and 1 if the company has foreign offices.

```
dom_int_data<-init_TAA_data<-init_TAA_data[c("Q24_box1","Q24_box2","Q24_box3","Q24_box4")]
dom_int<-ifelse(is.na(dom_int_data[,4]),0,1) #Create binary variable
init_TAA_data$Domestic_International<-dom_int
NA_index=apply(is.na(init_TAA_data[c("Q24_box1","Q24_box2","Q24_box3","Q24_box4")]),1,all) #Determine w
```

```r
init_TAA_data$Domestic_International[NA_index]=NA #Assign missingness pattern
```

Data Imputation (Xinyu)

Multivariate Regression Analysis (Qiang)

Regression Trees (Drew)

Linting, Mariëlle, Jacqueline J Meulman, Patrick JF Groenen, and Anita J van der Koojj. 2007. "Nonlinear Principal Components Analysis: Introduction and Application." *Psychological Methods* 12 (3): 336.