1-Regularized Multivariate

Mohammed Alrezq

2023-11-23

#install.packages("seminr")  
library(seminr)

## Warning: package 'seminr' was built under R version 4.2.3

library(tidyverse)

## Warning: package 'tibble' was built under R version 4.2.3

## Warning: package 'dplyr' was built under R version 4.2.3

library(psych)

## Warning: package 'psych' was built under R version 4.2.3

library(lavaan)

## Warning: package 'lavaan' was built under R version 4.2.3

library(semPlot)  
library(regsem)

## Warning: package 'regsem' was built under R version 4.2.3

## Warning: package 'Rcpp' was built under R version 4.2.3

## Warning: package 'Rsolnp' was built under R version 4.2.3

library("here")

## Warning: package 'here' was built under R version 4.2.3

library(corrplot)# to plot the correlation matrix   
library(glmnet)

## Warning: package 'glmnet' was built under R version 4.2.3

## Warning: package 'Matrix' was built under R version 4.2.3

# Download data

sem\_plsData <- read.csv("sem\_plsData.csv")  
# remove extra column   
sem\_plsData <- sem\_plsData |>  
 select(-1)

# Adding itmes for outcoems variabels by averaging

names(sem\_plsData)

## [1] "IMPL\_SUCC\_1" "BUDG\_COMP\_2" "TASK\_COMP\_3" "STAKEH\_ACC\_4" "DEPLOY\_6"   
## [6] "ACT\_INTG\_7" "EMPL\_SKI\_8" "EMPL\_SAT\_9" "EMPL\_BEL\_10" "BUSI\_IMP\_11"   
## [11] "ORG\_IMP\_12" "DEC\_IMP\_13" "LD\_SUP\_1" "EMP\_ATT\_3" "RES\_AVL\_4"   
## [16] "LD\_INV\_5" "PRJ\_SEL\_6" "DEF\_MEA\_7" "LD\_STL\_9" "DAT\_QLT\_11"   
## [21] "EMP\_ENG\_12" "TOOL\_AP\_13" "CONTX\_CUS\_14" "EFT\_REQ\_15" "TRA\_EDUC\_18"   
## [26] "WKF\_SKL\_19" "EXP\_CON\_20" "UND\_CNG\_22" "PAT\_RES\_24" "CUS\_IDEF\_26"

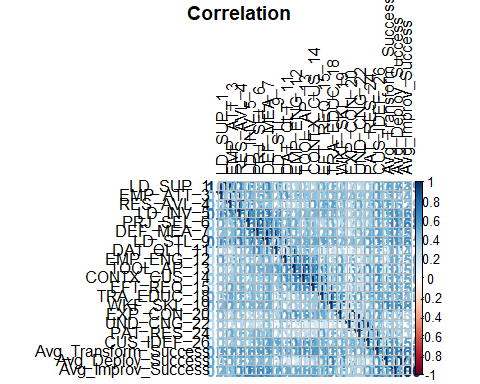
# Group the three outcomes and take their means   
sem\_plsData <- sem\_plsData |>   
 rowwise() |>   
 mutate(Avg\_Transform\_Success = mean(c\_across(c(IMPL\_SUCC\_1, EMPL\_SAT\_9, ORG\_IMP\_12, STAKEH\_ACC\_4, EMPL\_SKI\_8, DEC\_IMP\_13, BUSI\_IMP\_11, EMPL\_BEL\_10))),  
 Avg\_Deploy\_Success = mean(c\_across(c(DEPLOY\_6, ACT\_INTG\_7))),  
 Avg\_Improv\_Success = mean(c\_across(c(BUDG\_COMP\_2, TASK\_COMP\_3)))  
 )  
  
# Removes the itmes related to outcomes   
sem\_plsData <- sem\_plsData |>   
 select(-c(IMPL\_SUCC\_1, BUDG\_COMP\_2, TASK\_COMP\_3, STAKEH\_ACC\_4, DEPLOY\_6, ACT\_INTG\_7, EMPL\_SKI\_8, EMPL\_SAT\_9, EMPL\_BEL\_10, BUSI\_IMP\_11, ORG\_IMP\_12, DEC\_IMP\_13))

# Check for correlation

cor\_matrix <- round(cor(sem\_plsData, method = "spearman"), 2)  
#write.csv(cor\_matrix\_DVs, "DVs\_correlation\_Sperman.csv")  
# Plotting the correlation matrix  
corrplot(cor\_matrix,   
 method = "number",   
 #order = "hclust", # arranges the variables in a way that those with the highest similarities are closer to each other.   
 tl.col = "black",   
 #tl.srt = 45,   
 addCoef.col = "black",  
 title = "Correlation",  
 line = -1)

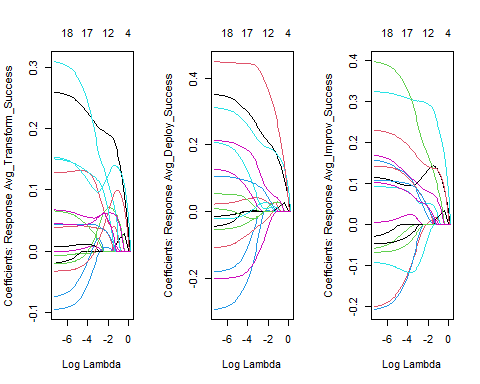
## Warning in text.default(pos.xlabel[, 1], pos.xlabel[, 2], newcolnames, srt =  
## tl.srt, : "line" is not a graphical parameter

## Warning in text.default(pos.ylabel[, 1], pos.ylabel[, 2], newrownames, col =  
## tl.col, : "line" is not a graphical parameter



#Note: overall correlation is okay less than 70%

#names(sem\_plsData)  
Y\_Var <- sem\_plsData[, c(19: 21)]  
X\_Var <- sem\_plsData[ ,c(1:18)]  
  
mutliv\_lasso <- glmnet(   
 as.matrix(X\_Var),  
 as.matrix(Y\_Var),  
 alpha = 1,  
 family = "mgaussian",  
 nlambda = 100)  
  
par(mfrow = c(1,3))  
plot(mutliv\_lasso, xvar = "lambda")



# Cross validation to determine optimal lambda   
mutliv\_lasso\_cv <- cv.glmnet(as.matrix(X\_Var),  
 as.matrix(Y\_Var),  
 family = "mgaussian",# mgaussian is used for multivariate outcomes   
 nfolds = 10,  
 type.measure = "mse")  
# optimla (min) lmbda   
mutliv\_lasso\_cv$lambda.min

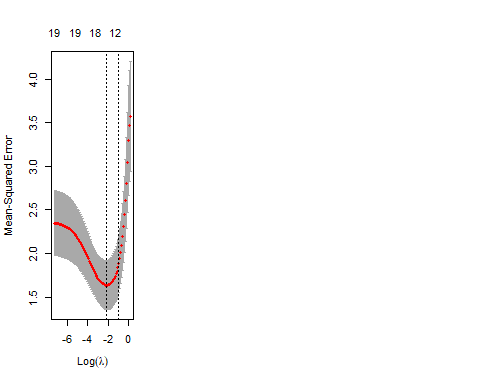
## [1] 0.1130907

mutliv\_lasso\_cv$lambda.1se

## [1] 0.379035

# plot   
#par(mfrow = c(1,3))  
plot(mutliv\_lasso\_cv)# Finla  
  
# Finla model with optimal lmbda   
mutliv\_lasso\_optimla <- glmnet(   
 as.matrix(X\_Var),  
 as.matrix(Y\_Var),  
 alpha = 1, # lasso   
 family = "mgaussian",  
 lambda = mutliv\_lasso\_cv$lambda.min)  
  
coef(mutliv\_lasso\_optimla)

## $Avg\_Transform\_Success  
## 19 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## 0.748340370  
## LD\_SUP\_1 0.192884896  
## EMP\_ATT\_3 0.054060566  
## RES\_AVL\_4 .   
## LD\_INV\_5 0.006319923  
## PRJ\_SEL\_6 0.116456353  
## DEF\_MEA\_7 .   
## LD\_STL\_9 .   
## DAT\_QLT\_11 0.031367888  
## EMP\_ENG\_12 0.068323473  
## TOOL\_AP\_13 .   
## CONTX\_CUS\_14 0.120949457  
## EFT\_REQ\_15 0.062056359  
## TRA\_EDUC\_18 .   
## WKF\_SKL\_19 0.054319289  
## EXP\_CON\_20 .   
## UND\_CNG\_22 0.031446047  
## PAT\_RES\_24 0.079733548  
## CUS\_IDEF\_26 0.056819632  
##   
## $Avg\_Deploy\_Success  
## 19 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## 0.179958747  
## LD\_SUP\_1 0.215719260  
## EMP\_ATT\_3 0.030394637  
## RES\_AVL\_4 .   
## LD\_INV\_5 0.003051092  
## PRJ\_SEL\_6 0.193724641  
## DEF\_MEA\_7 .   
## LD\_STL\_9 .   
## DAT\_QLT\_11 -0.022756133  
## EMP\_ENG\_12 0.008225323  
## TOOL\_AP\_13 .   
## CONTX\_CUS\_14 0.009961197  
## EFT\_REQ\_15 0.071356165  
## TRA\_EDUC\_18 .   
## WKF\_SKL\_19 0.378899675  
## EXP\_CON\_20 .   
## UND\_CNG\_22 0.051971962  
## PAT\_RES\_24 0.027117745  
## CUS\_IDEF\_26 -0.062229101  
##   
## $Avg\_Improv\_Success  
## 19 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## 0.0425868973  
## LD\_SUP\_1 0.1286112344  
## EMP\_ATT\_3 0.0005517563  
## RES\_AVL\_4 .   
## LD\_INV\_5 0.0263325542  
## PRJ\_SEL\_6 0.2959052843  
## DEF\_MEA\_7 .   
## LD\_STL\_9 .   
## DAT\_QLT\_11 0.0509678860  
## EMP\_ENG\_12 0.1787489246  
## TOOL\_AP\_13 .   
## CONTX\_CUS\_14 0.0520610256  
## EFT\_REQ\_15 0.0824350699  
## TRA\_EDUC\_18 .   
## WKF\_SKL\_19 0.1441503023  
## EXP\_CON\_20 .   
## UND\_CNG\_22 0.0503902582  
## PAT\_RES\_24 -0.0432571431  
## CUS\_IDEF\_26 0.0478887216



# 

# Transformational Success

names(sem\_plsData)

## [1] "LD\_SUP\_1" "EMP\_ATT\_3" "RES\_AVL\_4"   
## [4] "LD\_INV\_5" "PRJ\_SEL\_6" "DEF\_MEA\_7"   
## [7] "LD\_STL\_9" "DAT\_QLT\_11" "EMP\_ENG\_12"   
## [10] "TOOL\_AP\_13" "CONTX\_CUS\_14" "EFT\_REQ\_15"   
## [13] "TRA\_EDUC\_18" "WKF\_SKL\_19" "EXP\_CON\_20"   
## [16] "UND\_CNG\_22" "PAT\_RES\_24" "CUS\_IDEF\_26"   
## [19] "Avg\_Transform\_Success" "Avg\_Deploy\_Success" "Avg\_Improv\_Success"

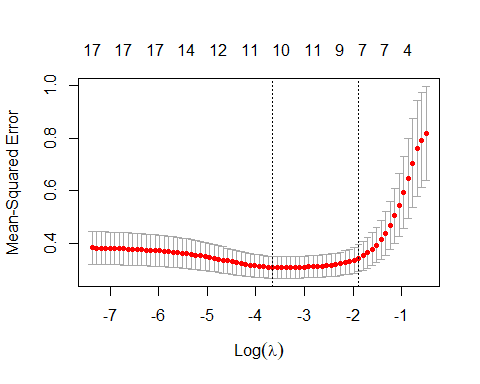
TS\_Y <- sem\_plsData[, 19]  
  
# Cross validation to determine optimal lambda   
TS\_lasso\_cv <- cv.glmnet(as.matrix(X\_Var),  
 as.matrix(TS\_Y),  
 family = "gaussian",# used for single outcome   
 nfolds = 10,  
 type.measure = "mse")  
# optimla (min) lmbda   
TS\_lasso\_cv$lambda.min

## [1] 0.02590946

TS\_lasso\_cv$lambda.1se

## [1] 0.1517522

# plot   
#par(mfrow = c(1,3))  
plot(TS\_lasso\_cv)# Finla



# Finla model with optimal lmbda   
TS\_lasso\_optimla <- glmnet(   
 as.matrix(X\_Var),  
 as.matrix(TS\_Y),  
 alpha = 1, # lasso   
 family = "gaussian",  
 lambda = TS\_lasso\_cv$lambda.min)  
  
coef(TS\_lasso\_optimla)

## 19 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## (Intercept) 0.37179213  
## LD\_SUP\_1 0.20320453  
## EMP\_ATT\_3 0.14930593  
## RES\_AVL\_4 .   
## LD\_INV\_5 .   
## PRJ\_SEL\_6 0.08243828  
## DEF\_MEA\_7 .   
## LD\_STL\_9 .   
## DAT\_QLT\_11 0.02963411  
## EMP\_ENG\_12 .   
## TOOL\_AP\_13 .   
## CONTX\_CUS\_14 0.23753815  
## EFT\_REQ\_15 0.02873174  
## TRA\_EDUC\_18 .   
## WKF\_SKL\_19 .   
## EXP\_CON\_20 0.02881962  
## UND\_CNG\_22 0.03039896  
## PAT\_RES\_24 0.12642027  
## CUS\_IDEF\_26 0.04335873

# Deployment Success

names(sem\_plsData)

## [1] "LD\_SUP\_1" "EMP\_ATT\_3" "RES\_AVL\_4"   
## [4] "LD\_INV\_5" "PRJ\_SEL\_6" "DEF\_MEA\_7"   
## [7] "LD\_STL\_9" "DAT\_QLT\_11" "EMP\_ENG\_12"   
## [10] "TOOL\_AP\_13" "CONTX\_CUS\_14" "EFT\_REQ\_15"   
## [13] "TRA\_EDUC\_18" "WKF\_SKL\_19" "EXP\_CON\_20"   
## [16] "UND\_CNG\_22" "PAT\_RES\_24" "CUS\_IDEF\_26"   
## [19] "Avg\_Transform\_Success" "Avg\_Deploy\_Success" "Avg\_Improv\_Success"

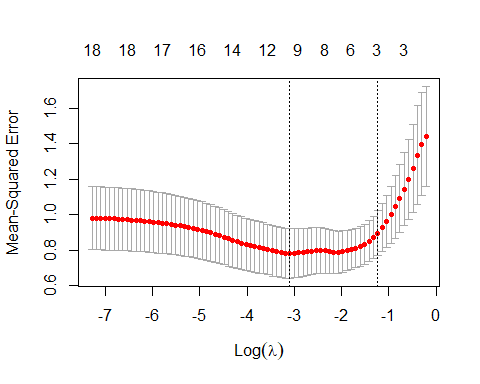
DS\_Y <- sem\_plsData[, 20]  
  
# Cross validation to determine optimal lambda   
DS\_lasso\_cv <- cv.glmnet(as.matrix(X\_Var),  
 as.matrix(DS\_Y),  
 family = "gaussian",# used for single outcome   
 nfolds = 10,  
 type.measure = "mse")  
# optimla (min) lmbda   
DS\_lasso\_cv$lambda.min

## [1] 0.04533421

DS\_lasso\_cv$lambda.1se

## [1] 0.2914116

# plot   
#par(mfrow = c(1,3))  
plot(DS\_lasso\_cv)# Finla



# Finla model with optimal lmbda   
DS\_lasso\_optimla <- glmnet(   
 as.matrix(X\_Var),  
 as.matrix(DS\_Y),  
 alpha = 1, # lasso   
 family = "gaussian",  
 lambda = DS\_lasso\_cv$lambda.min)  
  
coef(DS\_lasso\_optimla)

## 19 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## (Intercept) 0.16547377  
## LD\_SUP\_1 0.24175069  
## EMP\_ATT\_3 0.00713190  
## RES\_AVL\_4 0.01154336  
## LD\_INV\_5 .   
## PRJ\_SEL\_6 0.18242581  
## DEF\_MEA\_7 .   
## LD\_STL\_9 .   
## DAT\_QLT\_11 .   
## EMP\_ENG\_12 .   
## TOOL\_AP\_13 -0.06174686  
## CONTX\_CUS\_14 .   
## EFT\_REQ\_15 0.13089913  
## TRA\_EDUC\_18 .   
## WKF\_SKL\_19 0.45529715  
## EXP\_CON\_20 .   
## UND\_CNG\_22 0.08526264  
## PAT\_RES\_24 .   
## CUS\_IDEF\_26 -0.13175624

# Improvement Project Success

names(sem\_plsData)

## [1] "LD\_SUP\_1" "EMP\_ATT\_3" "RES\_AVL\_4"   
## [4] "LD\_INV\_5" "PRJ\_SEL\_6" "DEF\_MEA\_7"   
## [7] "LD\_STL\_9" "DAT\_QLT\_11" "EMP\_ENG\_12"   
## [10] "TOOL\_AP\_13" "CONTX\_CUS\_14" "EFT\_REQ\_15"   
## [13] "TRA\_EDUC\_18" "WKF\_SKL\_19" "EXP\_CON\_20"   
## [16] "UND\_CNG\_22" "PAT\_RES\_24" "CUS\_IDEF\_26"   
## [19] "Avg\_Transform\_Success" "Avg\_Deploy\_Success" "Avg\_Improv\_Success"

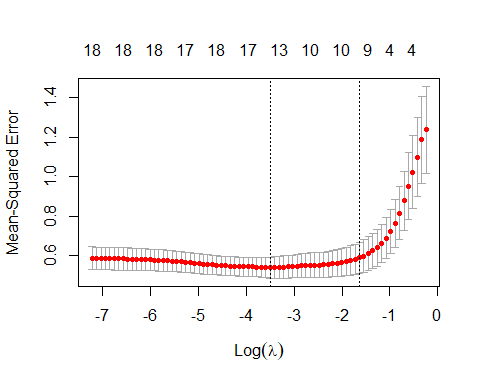
IPS\_Y <- sem\_plsData[, 21]  
  
# Cross validation to determine optimal lambda   
IPS\_lasso\_cv <- cv.glmnet(as.matrix(X\_Var),  
 as.matrix(IPS\_Y),  
 family = "gaussian",# used for single outcome   
 nfolds = 10,  
 type.measure = "mse")  
# optimla (min) lmbda   
IPS\_lasso\_cv$lambda.min

## [1] 0.030297

IPS\_lasso\_cv$lambda.1se

## [1] 0.1947513

# plot   
#par(mfrow = c(1,3))  
plot(IPS\_lasso\_cv)# Finla



# Finla model with optimal lmbda   
IPS\_lasso\_optimla <- glmnet(   
 as.matrix(X\_Var),  
 as.matrix(IPS\_Y),  
 alpha = 1, # lasso   
 family = "gaussian",  
 lambda = IPS\_lasso\_cv$lambda.min)  
  
coef(IPS\_lasso\_optimla)

## 19 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## (Intercept) -0.27738721  
## LD\_SUP\_1 0.05043818  
## EMP\_ATT\_3 -0.03643762  
## RES\_AVL\_4 .   
## LD\_INV\_5 0.13694535  
## PRJ\_SEL\_6 0.26865673  
## DEF\_MEA\_7 .   
## LD\_STL\_9 0.01922011  
## DAT\_QLT\_11 0.11253966  
## EMP\_ENG\_12 0.28246430  
## TOOL\_AP\_13 .   
## CONTX\_CUS\_14 .   
## EFT\_REQ\_15 0.09074697  
## TRA\_EDUC\_18 .   
## WKF\_SKL\_19 0.15669114  
## EXP\_CON\_20 -0.02506850  
## UND\_CNG\_22 0.09150461  
## PAT\_RES\_24 -0.09977586  
## CUS\_IDEF\_26 0.03694313