midterm_report.R

boburjonbahtiyorov

2022-10-26

```
### Correlation analysis
library(psych)
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(ltm)
```

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

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

## Loading required package: msm

## Loading required package: polycor

##
## Attaching package: 'polycor'

## The following object is masked from 'package:psych':
##
##     polyserial

##
## Attaching package: 'ltm'
```

```
## The following object is masked from 'package:psych':
##
##     factor.scores
```

```
library(tidyr)


df <- read.csv("/Users/boburjonbahtiyorov/Documents/University/AJOU/Statistical methods/Midterm exam /JI
str(df)
```

```
## 'data.frame':    331 obs. of  32 variables:
##  $ FIRMDUM1       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FIRMDUM2       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FIRMDUM3       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FIRMDUM4       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FIRMDUM5       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FIRMDUM6       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FIRMDUM7       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ InFreqGT1Deg   : num  42.5 60 72.5 37.5 65 52.5 42.5 30 47.5 47.5 ...
##  $ InCloseGT1Deg  : num  92.5 97.5 100 97.5 100 100 97.5 92.5 100 90 ...
##  $ RCP1           : int  5 7 6 6 6 7 7 7 7 6 ...
##  $ RCP2           : int  5 7 7 6 6 7 7 6 7 6 ...
##  $ RCP3           : int  5 7 6 6 6 7 7 6 7 6 ...
##  $ RWD1           : int  3 5 1 2 2 5 5 4 5 4 ...
##  $ RWD2           : int  4 5 1 3 2 4 5 5 4 4 ...
##  $ RWD3           : int  4 5 1 5 4 4 7 5 6 4 ...
##  $ RWD4           : int  4 5 1 5 3 4 7 5 6 4 ...
##  $ KTF1           : int  3 5 5 3 6 6 5 7 5 4 ...
##  $ KTF2           : int  4 6 4 3 4 6 5 4 4 4 ...
##  $ KTF3           : int  4 5 7 5 6 7 7 7 7 6 ...
##  $ KTF4           : int  5 5 7 5 6 7 7 7 7 6 ...
##  $ KTF5           : int  5 6 7 5 6 7 7 7 7 6 ...
##  $ KTF6           : int  4 6 7 4 6 7 7 6 6 5 ...
##  $ TNR            : int  97 181 110 31 73 73 12 24 49 86 ...
##  $ CMPTNR         : int  97 181 25 31 73 73 12 24 49 62 ...
##  $ JOBTNR         : int  97 181 75 12 65 38 12 12 25 48 ...
##  $ SEX            : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ AGE            : int  34 40 33 30 32 32 26 30 30 34 ...
##  $ RCP            : num  5 7 6.33 6 6 ...
##  $ RWD            : num  3.75 5 1 3.75 2.75 4.25 6 4.75 5.25 4 ...
##  $ KTF            : num  4.17 5.5 6.17 4.17 5.67 ...
##  $ OutGT1OutExpAvg: num  1.62 4.11 2.95 2.36 1.62 ...
##  $ Centrality     : num  67.5 78.8 86.2 67.5 82.5 ...
```

```
colnames(df)
```

```
##  [1] "FIRMDUM1"        "FIRMDUM2"        "FIRMDUM3"        "FIRMDUM4"
##  [5] "FIRMDUM5"        "FIRMDUM6"        "FIRMDUM7"        "InFreqGT1Deg"
##  [9] "InCloseGT1Deg"   "RCP1"            "RCP2"            "RCP3"
## [13] "RWD1"            "RWD2"            "RWD3"            "RWD4"
## [17] "KTF1"            "KTF2"            "KTF3"            "KTF4"
## [21] "KTF5"            "KTF6"            "TNR"             "CMPTNR"
## [25] "JOBTNR"          "SEX"             "AGE"             "RCP"
## [29] "RWD"             "KTF"             "OutGT1OutExpAvg" "Centrality"
```

```r
class(df)
```

```
## [1] "data.frame"
```

```r
df1 <- subset(df, select = c(Centrality, RCP,CMPTNR,OutGT1OutExpAvg,RWD,KTF ))

# Descriptive statistics and results of reliability test
describe(df1)
```

```
##                  vars   n  mean    sd median trimmed   mad   min max  range
## Centrality          1 331 70.68 15.13  70.83   71.08 14.85 12.24 100  87.76
## RCP                 2 331  6.25  0.82   6.33    6.38  0.99  2.67   7   4.33
## CMPTNR              3 331 51.80 51.56  32.00   42.69 28.17  1.00 276 275.00
## OutGT1OutExpAvg     4 331  3.45  0.86   3.51    3.49  0.89  1.37   5   3.63
## RWD                 5 331  3.82  1.31   4.00    3.81  1.11  1.00   7   6.00
## KTF                 6 331  5.26  0.93   5.33    5.30  0.99  1.00   7   6.00
##                   skew kurtosis   se
## Centrality       -0.58     1.35 0.83
## RCP              -1.26     1.85 0.05
## CMPTNR            1.64     2.42 2.83
## OutGT1OutExpAvg -0.35    -0.53 0.05
## RWD              0.02    -0.19 0.07
## KTF              -0.52     0.68 0.05
```

```r
reward <- subset(df, select = c(RWD1,RWD2,RWD3,RWD4))
centrality <- subset(df, select = c(InFreqGT1Deg,InCloseGT1Deg))
reciprocity <- subset(df, select = c(RCP1,RCP2,RCP3))
knowledge_transfer <- subset(df, select = c(KTF1,KTF2,KTF3,KTF4,KTF5,KTF6))

# Omitting missing values
sum(is.na(knowledge_transfer))
```

```
## [1] 2
```

```r
sum(is.na(reward))
```

```
## [1] 3
```

```r
reward <- na.omit(reward)
knowledge_transfer <- na.omit(knowledge_transfer)

# Cronbach's a
# Cronbach's a of Reward construct
cronbach.alpha(reward)
```

```
##
## Cronbach's alpha for the 'reward' data-set
##
## Items: 4
## Sample units: 330
## alpha: 0.89
```

```r
# Cronbach's a of Centrality construct
cronbach.alpha(centrality)
```

```
##
## Cronbach's alpha for the 'centrality' data-set
##
## Items: 2
## Sample units: 331
## alpha: 0.774
```

```r
# Cronbach's a of Reciprocity construct
cronbach.alpha(reciprocity)
```

```
##
## Cronbach's alpha for the 'reciprocity' data-set
##
## Items: 3
## Sample units: 331
## alpha: 0.895
```

```r
# Cronbach's a of Knowledge Transfer construct
cronbach.alpha(knowledge_transfer)
```

```
##
## Cronbach's alpha for the 'knowledge_transfer' data-set
##
## Items: 6
## Sample units: 329
## alpha: 0.83
```

```r
# The reliability of the research constructs were analyzed through Cronbach's alpha
# All the constructs showed reliability ranging from 0.774 to 0.895 which is higher than
# suggested threshold value of 0.7



# Correlations between constructs

lowerCor(df1, digits=3, use="pairwise", method="pearson")
```

```
##                 Cntrl  RCP    CMPTN  OGT1O  RWD    KTF
## Centrality       1.000
## RCP              0.014  1.000
## CMPTNR           0.199  0.063  1.000
## OutGT1OutExpAvg -0.052  0.136 -0.014  1.000
## RWD              0.101  0.172 -0.020  0.000  1.000
## KTF              0.257  0.414  0.117  0.126  0.310  1.000
```

```r
# Multidisciplinary test showed high relativity among reciprocity and knowledge transfer
```

```r
# Factor Analysis
centrality_fa <- subset(df, select = c(InFreqGT1Deg,InCloseGT1Deg))
reciprocity_fa <- subset(df, select = c(RCP1,RCP2,RCP3))
tenure_fa <- subset(df,select=c(CMPTNR))
embedded_r_fa <- subset(df,select=c(OutGT1OutExpAvg))
reward_fa <- subset(df, select = c(RWD1,RWD2,RWD3,RWD4))

sum(is.na(reward_fa))
```

```
## [1] 3
```

```r
reward_fa <- na.omit(reward_fa)


df_fa <- subset(df, select = c(InFreqGT1Deg, InCloseGT1Deg,RCP1,RCP2,RCP3,
                               CMPTNR,OutGT1OutExpAvg,RWD1,RWD2,RWD3,RWD4 ))
df_fa <- na.omit(df_fa)



# Factor analysis with principal component analysis and varimax rotation
pca = psych::principal(df_fa, nfactors=5, rotate="varimax")
print(pca, digits = 3, sort = FALSE)
```

```
## Principal Components Analysis
## Call: psych::principal(r = df_fa, nfactors = 5, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                     RC1    RC2    RC3    RC5    RC4    h2      u2  com
## InFreqGT1Deg      0.066  0.054  0.901  0.114  0.064 0.836 0.16417 1.06
## InCloseGT1Deg     0.036 -0.052  0.903  0.025 -0.103 0.831 0.16919 1.04
## RCP1              0.040  0.916  0.079  0.001  0.019 0.847 0.15306 1.02
## RCP2              0.115  0.895 -0.074  0.062  0.044 0.826 0.17404 1.06
## RCP3              0.098  0.901 -0.004 -0.001  0.058 0.825 0.17484 1.03
## CMPTNR           -0.023  0.044  0.127  0.982 -0.009 0.983 0.01742 1.04
## OutGT1OutExpAvg -0.001  0.087 -0.035 -0.009  0.991 0.991 0.00887 1.02
## RWD1              0.834  0.012  0.122 -0.120 -0.078 0.731 0.26912 1.10
## RWD2              0.914  0.072 -0.003 -0.024 -0.028 0.842 0.15816 1.02
## RWD3              0.839  0.096 -0.019  0.141  0.035 0.734 0.26601 1.09
## RWD4              0.869  0.117  0.044 -0.026  0.074 0.777 0.22334 1.06
##
##                        RC1   RC2   RC3   RC5   RC4
## SS loadings          3.020 2.496 1.673 1.017 1.016
## Proportion Var       0.275 0.227 0.152 0.092 0.092
## Cumulative Var       0.275 0.501 0.654 0.746 0.838
## Proportion Explained 0.327 0.271 0.181 0.110 0.110
## Cumulative Proportion 0.327 0.598 0.780 0.890 1.000
##
## Mean item complexity =  1
## Test of the hypothesis that 5 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.043
##  with the empirical chi square  67.543  with prob <  1.32e-10
##
```

```
## Fit based upon off diagonal values = 0.981
```

```
# Regression
colnames(df1)
```

```
## [1] "Centrality"      "RCP"             "CMPTNR"          "OutGT1OutExpAvg"
## [5] "RWD"             "KTF"
```

```
# all variables regressed on Knowledge Transfer
regression_KTF <- lm(KTF ~ Centrality + RCP + CMPTNR + OutGT1OutExpAvg + RWD, data = df1)
summary(regression_KTF)
```

```
##
## Call:
## lm(formula = KTF ~ Centrality + RCP + CMPTNR + OutGT1OutExpAvg +
##     RWD, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.63497 -0.46390  0.04569  0.53788  2.49176
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.7759813  0.4150753   1.869   0.0625 .
## Centrality      0.0136448  0.0029390   4.643 5.00e-06 ***
## RCP             0.4026693  0.0541594   7.435 9.37e-13 ***
## CMPTNR          0.0010128  0.0008592   1.179   0.2394
## OutGT1OutExpAvg 0.0965203  0.0506351   1.906   0.0575 .
## RWD             0.1614376  0.0337953   4.777 2.70e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7861 on 325 degrees of freedom
## Multiple R-squared:  0.2933, Adjusted R-squared:  0.2824
## F-statistic: 26.97 on 5 and 325 DF,  p-value: < 2.2e-16
```

```
# All variables are regressed on Embedded Resources
regression_EMB <- lm(OutGT1OutExpAvg~ Centrality + RCP + CMPTNR + KTF + RWD, data = df1)
summary(regression_EMB)
```

```
##
## Call:
## lm(formula = OutGT1OutExpAvg ~ Centrality + RCP + CMPTNR + KTF +
##     RWD, data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2316 -0.5122  0.1373  0.6083  1.5348
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.6654252  0.4298990   6.200 1.71e-09 ***
```

```
## Centrality  -0.0043505  0.0032975  -1.319   0.1880
## RCP          0.1003486  0.0635795   1.578   0.1155
## CMPTNR      -0.0003346  0.0009379  -0.357   0.7215
## KTF          0.1145522  0.0600947   1.906   0.0575 .
## RWD         -0.0311484  0.0380484  -0.819   0.4136
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8564 on 325 degrees of freedom
## Multiple R-squared:  0.03268,    Adjusted R-squared:  0.0178
## F-statistic: 2.196 on 5 and 325 DF,  p-value: 0.05444
```

```
# regression1 reciprocity regressed on embedded resources
colnames(df1)
```

```
## [1] "Centrality"      "RCP"             "CMPTNR"          "OutGT1OutExpAvg"
## [5] "RWD"             "KTF"
```

```
df1 <- na.omit(df1)
regression1 <- lm(OutGT1OutExpAvg ~ RCP, data = df1)
nrow(model.frame(regression1))
```

```
## [1] 331
```

```
summary(regression1)
```

```
##
## Call:
## lm(formula = OutGT1OutExpAvg ~ RCP, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.09914 -0.50638  0.08142  0.63212  1.58483
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.55742    0.36258   7.053 1.04e-11 ***
## RCP          0.14302    0.05749   2.488   0.0133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8574 on 329 degrees of freedom
## Multiple R-squared:  0.01847,    Adjusted R-squared:  0.01548
## F-statistic:  6.19 on 1 and 329 DF,  p-value: 0.01334
```

```
# knowledge transfer  regressed on embedded resources
regression2 <- lm(OutGT1OutExpAvg ~ KTF, data = df1)
nrow(model.frame(regression2))
```

```
## [1] 331
```

```
summary(regression2)
```

```
##
## Call:
## lm(formula = OutGT1OutExpAvg ~ KTF, data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1464 -0.5745  0.1124  0.5951  1.6960
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.83473    0.27204  10.420   <2e-16 ***
## KTF          0.11731    0.05093   2.303   0.0219 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8585 on 329 degrees of freedom
## Multiple R-squared:  0.01587,    Adjusted R-squared:  0.01288
## F-statistic: 5.306 on 1 and 329 DF,  p-value: 0.02188
```

```
# knowledge transfer regressed on both embedded resources and reciprocity
regression2 <- lm(OutGT1OutExpAvg ~ KTF + RCP, data = df1)
summary(regression2)
```

```
##
## Call:
## lm(formula = OutGT1OutExpAvg ~ KTF + RCP, data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1550 -0.5398  0.1077  0.6220  1.6384
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.37472    0.38473   6.172 1.98e-09 ***
## KTF          0.07833    0.05580   1.404   0.1613
## RCP          0.10634    0.06307   1.686   0.0927 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8561 on 328 degrees of freedom
## Multiple R-squared:  0.02433,    Adjusted R-squared:  0.01838
## F-statistic: 4.089 on 2 and 328 DF,  p-value: 0.01761
```

```
# centrality and reciprocity on knowledge transfer
regression3 <- lm(KTF ~ Centrality + RCP, data = df1)
summary(regression3)
```

```
##
## Call:
## lm(formula = KTF ~ Centrality + RCP, data = df1)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.96660 -0.49887  0.06912  0.61382  2.23972
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.267354   0.400645   3.163  0.00171 **
## Centrality  0.015407   0.002963   5.200 3.51e-07 ***
## RCP         0.464386   0.054600   8.505 6.52e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8143 on 328 degrees of freedom
## Multiple R-squared:  0.2347, Adjusted R-squared:   0.23
## F-statistic:  50.3 on 2 and 328 DF,  p-value: < 2.2e-16
```

```
# Company Tenure influence on knowledge transfer
regression4 <- lm(OutGT1OutExpAvg ~ CMPTNR, data = df1)
summary(regression4)
```

```
##
## Call:
## lm(formula = OutGT1OutExpAvg ~ CMPTNR, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.09482 -0.54233  0.05899  0.64392  1.58319
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.4637765  0.0674739  51.335   <2e-16 ***
## CMPTNR      -0.0002302  0.0009238  -0.249    0.803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8654 on 329 degrees of freedom
## Multiple R-squared:  0.0001888,  Adjusted R-squared:  -0.00285
## F-statistic: 0.06211 on 1 and 329 DF,  p-value: 0.8033
```

```
# Embedded resources on knowledge transfer
regression5 <- lm(KTF ~ OutGT1OutExpAvg, data = df1)
summary(regression5)
```

```
##
## Call:
## lm(formula = KTF ~ OutGT1OutExpAvg, data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2248 -0.5858  0.0638  0.6567  1.8964
##
## Coefficients:
```

```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.79352    0.20898  22.938   <2e-16 ***
## OutGT1OutExpAvg  0.13529    0.05873   2.303   0.0219 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.922 on 329 degrees of freedom
## Multiple R-squared:  0.01587,    Adjusted R-squared:  0.01288
## F-statistic: 5.306 on 1 and 329 DF,  p-value: 0.02188
```

```
detach(package:psych, unload = TRUE)
```