REPORT ON BAROMETRO CAMTO 2018

INTRODUCTION

METHODS

Statistical assumptions in Likert scales

Explain missing data pattern and employed techniques

It is comprised by 3 Likert type scales (7 responde options)

How everything is calculated in the questionnaire.

For instance:

Efficacy1 =

Efficacy 2 =

RESULTS

Descriptives

Años de encuestas completados:

12 encuestas de 2015, 3 de 2016 y 24 de 2017

Completos: 12 casos

Casos totales 39

La moda del sector de las empresas es &: teconologia y telecomunicaciones)

ESCALA IDENTIFICACION

> nitemsID <- sum(length(6:12)); nitemsID

[1] 7

> dimID <- nitemsID\*ncases; dimID

[1] 273

> NAid <- sum(is.na(scaleID)); NAid

[1] 33

> percentageNAid <- NAid/dimID; percentageNAid

[1] 0.1208791

> cat("percentage of missing data in this scale is", percentageNAid\*100)

percentage of missing data in this scale is 12.08791

ESCALA FUERZA DE TRABAJO

> nitemsFT <- sum(length(13:27)); nitemsFT

[1] 15

> dimFT <- nitemsFT\*ncases; dimFT

[1] 585

> NAft <- sum(is.na(scaleFT)); NAft

[1] 111

> percentageNAft <- NAft/dimFT; percentageNAft

[1] 0.1897436

> cat("percentage of missing data in this scale is", percentageNAft\*100)

percentage of missing data in this scale is 18.97436

> as.matrix(round(a$mean))

[,1]

[1,] 20146

[2,] 20506

[3,] 39539

[4,] 0

[5,] 0

[6,] 0

[7,] 0

[8,] 1

[9,] 0

[10,] 39

[11,] 1

[12,] 2

[13,] 3

[14,] 2

[15,] 1

ESCALA EFICACIA

> nitemsEF <- sum(length(28:34)); nitemsEF

[1] 7

> dimEF <- nitemsEF\*ncases; dimEF

[1] 273

> NAef <- sum(is.na(scaleef)); NAef

[1] 2

> percentageNAef <- NAef/dimEF; percentageNAef

[1] 0.007326007

> cat("percentage of missing data in this scale is", percentageNAef\*100)

percentage of missing data in this scale is 0.7326007

ESCALA GENERAL

> nitemsgen <- sum(length(35:40)); nitemsgen

[1] 6

> dimGEN <- nitemsgen\*ncases; dimGEN

[1] 234

> NAgen <- sum(is.na(scalegen)); NAgen

[1] 42

> percentageNAgen <- NAgen/dimGEN; percentageNAgen

[1] 0.1794872

> cat("percentage of missing data in this scale is", percentageNAgen\*100)

percentage of missing data in this scale is 17.94872

ESCALA ESPECIFICA

> dimESP <- nitemsESP\*ncases; dimESP

[1] 975

> NAESP <- sum(is.na(scaleesp)); NAESP

[1] 102

> percentageNAESP <- NAESP/dimESP; percentageNAESP

[1] 0.1046154

> cat("percentage of missing data in this scale is", percentageNAESP\*100)

percentage of missing data in this scale is 10.46154

TOTAL ESCALAS

> # 2 MISSING DATA

> NAindata <- sum(is.na(baro[ , -(1:4)])); NAindata

[1] 290

> percentageNAbaro <- NAindata/prod(dim(baro[ , -(1:4)])); percentageNAbaro

[1] 0.1219

> cat("percentage of missing data is", percentageNAbaro\*100)

percentage of missing data is 12.19

SCALES PROPERTIES

> #Check unidimensionality

> fa.parallel(scaleef)

Parallel analysis suggests that the number of factors = 1 and the number of components = 1

> fa.parallel(scalegen)

Parallel analysis suggests that the number of factors = 1 and the number of components = 1

> fa.parallel(scaleesp)

Parallel analysis suggests that the number of factors = 2 and the number of components = 1

> fa(scaleesp, nfactors=2)

Factor Analysis using method = minres

Call: fa(r = scaleesp, nfactors = 2)

Warning: A Heywood case was detected.

Standardized loadings (pattern matrix) based upon correlation matrix

MR1 MR2 h2 u2 com

ESPE01 0.31 0.59 0.67 0.33 1.5

ESPE02 0.31 0.34 0.33 0.67 2.0

ESPE03 0.34 0.35 0.39 0.61 2.0

ESPE04 0.02 0.75 0.58 0.42 1.0

ESPE05 0.56 0.25 0.54 0.46 1.4

ESPE06 0.08 0.63 0.47 0.53 1.0

ESPE07 0.45 0.42 0.61 0.39 2.0

ESPE08 0.63 0.27 0.68 0.32 1.4

ESPE09 0.16 0.62 0.53 0.47 1.1

ESPE10 -0.04 0.88 0.74 0.26 1.0

ESPE11 -0.14 1.00 0.85 0.15 1.0

ESPE12 0.08 0.80 0.72 0.28 1.0

ESPE13 0.11 0.74 0.67 0.33 1.0

ESPE14 0.91 0.00 0.82 0.18 1.0

ESPE15 0.34 0.59 0.71 0.29 1.6

ESPE16 0.51 0.33 0.57 0.43 1.7

ESPE17 0.82 -0.03 0.64 0.36 1.0

ESPE18 0.69 0.23 0.73 0.27 1.2

ESPE19 0.55 0.30 0.60 0.40 1.5

ESPE20 0.79 0.08 0.71 0.29 1.0

ESPE21 0.63 0.26 0.66 0.34 1.3

ESPE22 0.66 -0.10 0.36 0.64 1.0

ESPE23 0.79 -0.14 0.51 0.49 1.1

ESPE24 0.87 -0.21 0.57 0.43 1.1

ESPE25 0.68 0.04 0.50 0.50 1.0

MR1 MR2

SS loadings 8.21 6.95

Proportion Var 0.33 0.28

Cumulative Var 0.33 0.61

Proportion Explained 0.54 0.46

Cumulative Proportion 0.54 1.00

With factor correlations of

MR1 MR2

MR1 1.00 0.62

MR2 0.62 1.00

Mean item complexity = 1.3

Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are 300 and the objective function was 37.3 with Chi Square of 1075.6

The degrees of freedom for the model are 251 and the objective function was 18.13

The root mean square of the residuals (RMSR) is 0.08

The df corrected root mean square of the residuals is 0.09

The harmonic number of observations is 39 with the empirical chi square 155.58 with prob < 1

The total number of observations was 39 with Likelihood Chi Square = 498.6 with prob < 1.6e-18

Tucker Lewis Index of factoring reliability = 0.592

RMSEA index = 0.214 and the 90 % confidence intervals are 0.14 NA

BIC = -420.95

Fit based upon off diagonal values = 0.98

Measures of factor score adequacy

MR1 MR2

Correlation of (regression) scores with factors 0.99 0.99

Multiple R square of scores with factors 0.98 0.97

Minimum correlation of possible factor scores 0.96 0.95

Warning message:

In fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, :

A loading greater than abs(1) was detected. Examine the loadings carefully.

> fa(scaleesp, nfactors=1)

Factor Analysis using method = minres

Call: fa(r = scaleesp, nfactors = 1)

Standardized loadings (pattern matrix) based upon correlation matrix

MR1 h2 u2 com

ESPE01 0.80 0.65 0.35 1

ESPE02 0.58 0.33 0.67 1

ESPE03 0.62 0.39 0.61 1

ESPE04 0.67 0.45 0.55 1

ESPE05 0.73 0.54 0.46 1

ESPE06 0.63 0.40 0.60 1

ESPE07 0.78 0.61 0.39 1

ESPE08 0.82 0.67 0.33 1

ESPE09 0.69 0.47 0.53 1

ESPE10 0.73 0.54 0.46 1

ESPE11 0.74 0.55 0.45 1

ESPE12 0.77 0.59 0.41 1

ESPE13 0.75 0.57 0.43 1

ESPE14 0.82 0.68 0.32 1

ESPE15 0.83 0.69 0.31 1

ESPE16 0.76 0.57 0.43 1

ESPE17 0.72 0.53 0.47 1

ESPE18 0.84 0.70 0.30 1

ESPE19 0.77 0.59 0.41 1

ESPE20 0.79 0.63 0.37 1

ESPE21 0.81 0.65 0.35 1

ESPE22 0.51 0.26 0.74 1

ESPE23 0.60 0.36 0.64 1

ESPE24 0.60 0.36 0.64 1

ESPE25 0.66 0.44 0.56 1

MR1

SS loadings 13.22

Proportion Var 0.53

Mean item complexity = 1

Test of the hypothesis that 1 factor is sufficient.

The degrees of freedom for the null model are 300 and the objective function was 37.3 with Chi Square of 1075.6

The degrees of freedom for the model are 275 and the objective function was 21.17

The root mean square of the residuals (RMSR) is 0.11

The df corrected root mean square of the residuals is 0.11

The harmonic number of observations is 39 with the empirical chi square 280.87 with prob < 0.39

The total number of observations was 39 with Likelihood Chi Square = 596.2 with prob < 8.3e-26

Tucker Lewis Index of factoring reliability = 0.533

RMSEA index = 0.225 and the 90 % confidence intervals are 0.156 NA

BIC = -411.28

Fit based upon off diagonal values = 0.96

Measures of factor score adequacy

MR1

Correlation of (regression) scores with factors 1.00

Multiple R square of scores with factors 1.00

Minimum correlation of possible factor scores 0.99

> #Check omega

> omega(scaleef,nfactors=1)

Omega Total 0.92

> omega(scalegen,nfactors=1)

Omega Total 0.92

> omega(scaleesp,nfactors=1)

Omega Total 0.97

DISCUSSION

While some quick fixes such as mean-substitution may be fine in some cases, such simple approaches usually introduce bias into the data, for instance, applying mean substitution leaves the mean unchanged (which is desirable) but decreases variance, which may be undesirable.

* MCAR: missing completely at random. This is the desirable scenario in case of missing data.
* MNAR: missing not at random. Missing not at random data is a more serious issue and in this case it might be wise to check the data gathering process further and try to understand why the information is missing. For instance, if most of the people in a survey did not answer a certain question, why did they do that? Was the question unclear?

Assuming data is MCAR, too much missing data can be a problem too. Usually a safe maximum threshold is 5% of the total for large datasets. If missing data for a certain feature or sample is more than 5% then you probably should leave that feature or sample out. We therefore check for features (columns) and samples (rows) where more than 5% of the data is missing using a simple function

LIMITATIONS

Small simple

Missing data

Low variability in certain subjects; thus low level of information provided

Assumptions about missing data needed to be made

There are several ways to compute work rotation; it changes depending on the Company.

Maybe it contains too many scales and thus too many ítems; making it difficult to comply with the whole questionnaire.

Maybe it can be sugested to display a Principal Components analysis to check ítem correlations and trim down the simple of ítems so the whole questionnaire might be easier to complete; thus enjoying a lesser degree of missing data.

REFERENCES

APPENDIX 1: R SCRIPT

APPENDIX 2: SURVEY IN SPANISH