Gregor Štiglic, Leona Cilar Budler, Roger Watson

15 Running a confirmatory factor analysis in R: a step-by-step tutorial

Abstract

Introduction: Confirmatory factor analysis is a frequently used data analysis technique in nursing research for the development of new measures and psychometric evaluation. R is an open-source statistical software for data analysis that allows a high level of reproducibility required for submission in more and more scientific journals in nursing. The aim of this chapter is to provide a step-by-step tutorial for conducting confirmatory factor analysis in the statistical program R.

Methods: Data collected from 938 students in Scotland who completed the Trait Emotional Intelligence Questionnaire Short form (TEIQue-SF) were used to conduct analyses.

Results: We demonstrated a reproducible way of conducting a correlation-based network visualization followed by confirmatory factor analysis to confirm the four-factor structure in the TEIQue-SF questionnaire.

Conclusion: This chapter focuses on the demonstration of an alternative statistical tool to the frequently used point-and-click software by providing detailed instructions in the form of reproducible code and data.

Keywords: nurses, nursing, R, data analysis, factor analysis, confirmatory factor analysis

15.1 Introduction

R is an open-source, crowd-funded, programming language for statistical computing and graphics [1]. R uses different packages for conducting various functions and analyses. The packages are created by statisticians across the world. Packages are available under a Creative Commons License and are regularly updated and rigorously tested before being made available via CRAN (The Comprehensive R Archive Network) [2, 3]. R can be used with various platforms including Linux, Mac, and Windows [4]. At the time of writing, version 4.0.2 (Taking Off Again) is available. R is not easy to use; it has a "steep learning curve" and can be hard for someone without computer programming experience. On the other hand, it is free (*caveat emptor* applies), and a lot of help is available online. The use of R can be facilitated by installing RStudio® (the basic package is free) and allows the user to store coding. Even

https://doi.org/10.1515/9783110786088-015

with this facility, users must workaround obscure error messages which appear frequently. Although R is a free, powerful, and flexible alternative, it is less familiar and less frequently used in nursing research [5].

Healthcare institutions must be prepared to invest in new, effective, and efficient treatments and interventions for patient benefit. Before the implementation of new interventions, research must be done using valid and reliable measurements [6]. Questionnaires used for the data collection need to be tested for reliability and validity. When developing and validating the questionnaire, psychometric tests must be conducted. These include, for example, parametric item response theory, factor analysis (FA), and reliability via classical test theory.

Exploratory analysis of data can be conducted to observe the correlations between items as an initial step to statistically more rigorous approaches. Correlation or covariance matrices are usually the simplest techniques that can be used in this step. While such an approach allows good visualization of multiple pairwise correlations, it is unable to capture the more complex correlations among multiple items. One of the novel approaches used in this field is correlation networks that aim, simultaneously, to capture the correlation and grouping of the items in the same visualization and were used in other fields like bioinformatics earlier [7]. However, the technique recently developed in multiple variants of psychological network visualization as a psychometric approach [8].

FA is a method for testing whether item covariance structure justifies the use of item scores to calculate (sub)scale scores [9]. The aim of FA is to reduce "the dimensionality of the original space and to give an interpretation to the new space, spanned by a reduced number of new dimensions which are supposed to underlie the old ones" [10]. Matsunaga [11] discusses the importance of FA and how it is often misunderstood. Although an increasing number of researchers are using these methods, evidence suggests that a large proportion of the research published still harbours ill-informed practices. Thus, it is important that researchers are aware of different FA methods. Confirmatory FA (CFA) is often used for developing a new measure, testing measurement invariance, construct validation, testing method effects, and psychometric evaluation. CFA is strongly related to exploratory FA (EFA), principal component analysis, and structural equation modelling (SEM) [12]. EFA is described as the orderly simplification of interrelated measures. It has been traditionally used to explore the possible underlying factor structure of a set of observed variables without imposing a preconceived structure on the outcome [13]. The result of performing EFA is identification of the underlying factor structure. On the other hand, CFA is used to verify the factor structure of a set of observed variables. It allows testing a relationship between observed variables and the existence of their underlying latent constructs. The researcher can use theoretical knowledge, empirical research, or both; can postulate the relationship pattern a priori; and can then test the hypothesis statistically [14]. Both the EFA and CFA are types of factor analyses based on the common factor model, which proposes that each observed response

is influenced partially by the underlying common factors and partially by the underlying unique factors [15]. For conducting CFA, researchers must have strong empirical or conceptual knowledge to guide factor model specification and evaluation [12]. CFA will be described in detail in the following sections.

In this tutorial, we introduce a short step-by-step tutorial on how to perform CFA for researchers using statistics and R programming in applied healthcare research. The corresponding R Markdown script is available as a supplementary material.

15.2 Methods

15.2.1 Design

This chapter represents a tutorial on using the R statistical software to perform CFA and the corresponding exploratory analysis.

15.2.2 Participants

We used the dataset originally from the study by Snowden et al. [16] to examine the construct validity of the Trait Emotional Intelligence Questionnaire Short form (TEIQue-SF). The TEIQue-SF is a short-form questionnaire which includes 30 items. It was developed by Petrides [17] to measure global trait emotional intelligence (trait EI). Petrides derived it from the larger 130-item-based TEIQue questionnaire by Freudenthaler et al. [18]. Items were selected based on their correlations with the corresponding total facet scores [19]. Responses to the TEIQue-SF items are made on a Likert-type scale (e.g., 1 meaning “strongly disagree” and 7 meaning “strongly agree”). The total scale score is derived by summing the score of each item (after reverse scoring of negative items) and is used to locate respondents on the latent trait continuum. A higher score represents a greater presence of the trait EI [20]. Four dimensions of trait EI are measured with the TEIQue-SF: well-being, sociability, self-control, and emotionality. Petrides [17] defines well-being as related feelings of a time based around achievements, self-regard, and expectations; self-control as regulating and having control over emotions, impulses, and stress; emotionality as an ability to perceive, express, and connect with emotions in self and others, which can be used in creating successful interpersonal relationships; and sociability as being socially assertive and aware, managing others’ emotions, and effectiveness in communication and participation in social situations. Sample characteristics are presented in Tab. 15.1 [16].

Tab. 15.1: Sample characteristics.

	Programme	Frequency	Per cent	Valid per cent	Cumulative per cent
Valid	Adult	586	62.5	62.5	62.5
	Mental health	124	13.2	13.2	75.8
	Learning disability	29	3.1	3.1	78.9
	Children	47	5.0	5.0	83.9
	Midwifery	83	8.8	8.9	92.7
	Computing	68	7.2	7.3	100.0
	Total	937	99.9	100.0	
Missing	System	1	0.1		
Total		938	100.0		

15.2.3 Data collection

Data were collected from 938 students of nursing, midwifery, and computer science in two Scottish universities. The dataset also contains information on the gender of the participants. The data and R source code are available at the Mendeley Data repository [21].

15.2.4 Data analysis

FA is a common method used to find a small set of unobserved variables (called latent variables, or factors) which can account for the covariance among a set of observed variables (called manifest variables) [22]. Latent variables are variables that are inferred, not directly observed, from other variables that are observed. The presence of latent variables can be detected by their effects on variables that are observable [23]. A manifest variable is a variable or factor that can be directly measured or observed. Commonly, CFA models are displayed as path diagrams in which squares represent observed variables and circles represent the latent concepts. Moreover, single-headed arrows are used to imply a direction of assumed causal influence, and double-headed arrows represent covariance between two latent variables [22].

Initially, we should justify the sample size which allows a robust enough CFA. Determining the appropriate sample size for CFA is a complex task; therefore, we refer the interested readers to read more on this topic in a paper by Kyriazos [24].

Also, before any FA, it is important to test the dataset for FA suitability. There are two methods for testing FA suitability: the Bartlett's Test of Sphericity and the Kaiser, Meyer, Olkin (KMO) test. Those tests are used to check whether a matrix is significantly different from an identity matrix. A significant Bartlett's test of sphericity states that FA may proceed if $p < 0.001$. In 1970, this method was modified by Kaiser as the Measure of Sampling Adequacy (MSA), and in 1974 by Kaiser and Rice [25]. The KMO statistic can vary from 0 to 1 and indicates the degree to which each variable in a set is predicted without error by the other variables. The KMO statistic indicates how much variance in your variables might be caused by underlying factors. High values, close to 1.0, indicate that a factor analysis may be appropriate for your data. CFA is based on the covariances among variances. These are susceptible to the effects of violations to the assumption of normality which can affect covariances. CFA models contain the parameters of factor variances and covariances (Ψ – psi), factor loadings (Λ – lambda), and error factor variances and covariances (Θϵ – theta–epsilon) [26].

Before running a model, the variables should be examined to check that there are no deviations from normality. The MVN package [27] provides various approaches to check for univariate or multivariate deviation from a normal distribution. In practice, there are many cases where normality will not be met, especially in surveys using Likert scales. In such cases, we can use a robust estimator as presented in Finney and DiStefano [28]. In this tutorial, we used maximum likelihood estimation with robust standard errors and a Satorra–Bentler scaled test statistic which can be used even in cases of deviations from the normality in most or all variables.

A simple exploratory visualization of the data can be performed to check for the correlations between the variables and to see the grouping of the variables in factors. It should be noted that correlation matrices produce different results from CFA when the relationships between the items are observed. Instead of using a correlation matrix, CFA uses a variance–covariance matrix on raw data to estimate input variance [12]. In this chapter, we present an example of correlation visualization using correlation network graphs [29]. Correlation graphs have been used to visualize relations and the grouping of variables at the same time. However, the most frequently used force-directed positioning of the network nodes representing variables is not easily interpretable and can be misleading when grouping of the variables is observed. Therefore, alternative options are available in the qgraph package [29]. In this chapter, we used multidimensional scaling to position nodes resulting in a more realistic grouping of nodes [8].

First, the factor loadings of the indicators (observed variables) that make up the latent construct need to be calculated. The standardized factor loading squared is the estimate of the amount of the variance of the indicator that is accounted for by the latent construct. Factor loadings of 0.4 or higher are acceptable [30]. The unique variance is not explained by the latent construct. We also need to check the convergent validity of the construct, which is indicated by high indicator loadings, which shows the strength of how well the indicators are theoretically similar. Most models

contain more than one factor. In that case, we need to run a CFA for all the model's latent constructs within one measurement model. Discriminant validity exists when no two constructs are highly correlated. If two constructs are highly correlated (> 0.85), we need to explore combining the constructs.

Many different approaches are proposed to assess the fit of the model to the data [31]. Some of the more popular fit statistics include the comparative fit index (CFI), the Tucker–Lewis index (TLI), and root mean square error of approximation (RMSEA). CFI is frequently used as it is known to perform well even when the sample size is small [32] and assumes that all latent variables are uncorrelated and compares the postulated model to the null model. The CFI values can range from 0.0 to 1.0, where values close to 1.0 represent a good fit. To avoid misspecified models it is generally advised for models to achieve CFI over 0.90 which represents a good fit with some authors arguing that the threshold should lie at 0.95 [33]. TLI measures a relative reduction in misfit per degree of freedom [34] with threshold values of 0.90 indicating adequate fit and 0.95 indicating good fit [35]. On the other hand, the RMSEA represents a so-called badness-of-fit measure resulting in lower values for a better fit. It measures discrepancy due to the approximation with values below 0.06 representing an acceptable model [36].

15.3 Results

In this section, we provide step-by-step instructions on how to run the CFA analysis.

15.3.1 Step 1: reading and checking the data

In the initial step, we read the data from the CSV file using the read.csv command in R. After reading the data, we checked whether the data were loaded by using a command str which prints the summary information on the structure of the data just loaded. This way it is possible to print the type and a few example values for each variable. In the case of the TEIQue-SF example data provided with this chapter, we observe that our data consists of 938 samples with 36 variables. It needs to be noted that as in most real-world datasets only a specific number of variables will be used in the CFA analysis. In our case, only 30 variables are TEIQue-SF scale variables. Items from the TEIQue-SF scale will be used in the CFA analysis. As mentioned, it is important that the researcher who is performing a CFA has knowledge about the scale and its basic properties. Thus, we removed items 3, 18, 14, and 29 that were considered as a "general" factor by Petrides and Furnham [19], and we were left with 26 items that should represent 4 factors: well-being, self-control, emotionality, and sociability. For visualization needed in the next step, we also import textual description for each item containing the questions for all 30 items.

In addition to the above-mentioned data reading and checking functions, the following packages need to be loaded in R: smacof (multidimensional scaling needed for exploratory analysis in our case), parameters (KMO and Bartlet's test), lavaan (CFA analysis), qgraph (visualization for exploratory analysis), and semplot (visualization of the relations between items and factors). The initial steps are shown in Box 15.1.

Box 15.1: Source code for exploratory analysis.

```
library(qgraph)
library(lavaan)
library(semPlot)
library(smacof)
library(parameters)
# Load the data from a csv file
data <- read.csv("db_TEIQ_CFA_SCO.csv", header = T)
# Print some information on the data we just read to check
# whether the data loaded properly
str(data)
# Separate TEIQue-SF values in a variable named teiq
teiq <- data[,2:31]
# Load item names
names <- readLines("items.txt")
# Create variable names for TEIQue-SF data (V1 – V30)
colnames(teiq) <- paste0("V", 1:30)
# Items 3, 18, 14 & 29 were omitted from the CFA because
# Petrides (2006) considered them a 'general' factor as
# they were not specifically associated with any particular factor.
# Additionally, item number 20 was omitted as suggested by
# Snowden et al. (2016)
teiq <- teiq[,-c(3, 18, 14, 29, 20)]
names <- names[-c(3, 18, 14, 29, 20)]
# Rearange columns and related questions to match four subscales
items <- c("V30","V15","V19","V24","V27","V21","V9","V6",
"V12","V5","V28","V13","V16",
"V7","V10","V22","V25","V8","V4","V2",
"V11","V26","V17","V1","V23")
pos <- match(items, colnames(teiq))
teiq <- teiq[,items]
names <- names[pos]
```

15.3.2 Step 2: exploratory data analysis

The Bartlett's test of sphericity and the KMO tests were conducted as shown in Box 15.2 to test if the dataset is suitable for FA. Moreover, variables were examined

to check for deviations from normality using univariate (Shapiro–Wilk) and multivariate (Henze–Zirkler) tests of normality.

The KMO MSA suggests that the data are appropriate for FA (KMO = 0.88). Bartlett's test of sphericity suggests that there is sufficient significant correlation in the data for FA ($\chi^2(300) = 4{,}841.03$, $p < 0.001$). On the other hand, none of the items is normally distributed and consequently, the multivariate normality test shows the same. Since we are using Likert-scale items, this is not unusual, but it requires a different approach to the CFA. In our case, a robust estimator following Finney and DiStefano [28] was used to perform CFA on non-normally distributed data. More specifically, we used maximum likelihood estimation with robust standard errors and a Satorra–Bentler scaled test statistic.

Box 15.2: Source code for basic CFA data assessment.

```
# KMO and Bartlet's test
check_factorstructure(teiq)
# Univariate (Shapiro-Wilk) and multivariate (Henze-Zirkler) test of normal distribution
result <- mvn(data = teiq, mvnTest = "hz", univariateTest = "SW", desc = TRUE)
result$univariateNormality
result$multivariateNormality
```

Next, we used a simple exploratory visualization of the data to check the correlations between the variables and potentially already see the grouping of variables into four factors. This can be done using the R command qgraph (Fig. 15.1 and Box 15.3).

Box 15.3: Visualization of the data using correlation graph.

```
# Define a vector of item groups belonging to subscales
groups <- c(rep('Confidence', 8), rep('Connection', 5),
rep('Uncertainty', 7), rep('Empathy', 5))
group_col <- c("#72CF53", "#53B0CF", "#FFB026", "#ED3939")
# Covariance matrix
corMat <- cov(teiq, use = "pairwise.complete.obs")
# Multidimensional scaling based positioning of the nodes in the network
dissimilarity <- sim2diss(cor(teiq))
mdsModel <- mds(dissimilarity)
head(round(mdsModel$conf, 2))
png(filename = "figure2.png", type = "cairo", height = 6, width = 12, units = 'in', res = 300)
qgraph(corMat, graph = "cor", sampleSize = nrow(data),
layout = mdsModel$conf, color = group_col,vsize = 4, esize = 4,
border.width = 2, border.color = "black", groups = groups,
nodeNames = names, legend = TRUE, legend.mode = "style1",
legend.cex = .36, theme = "colorblind",
threshold = "bonferroni",
minimum = "sig", alpha = 0.05)
dev.off()
```

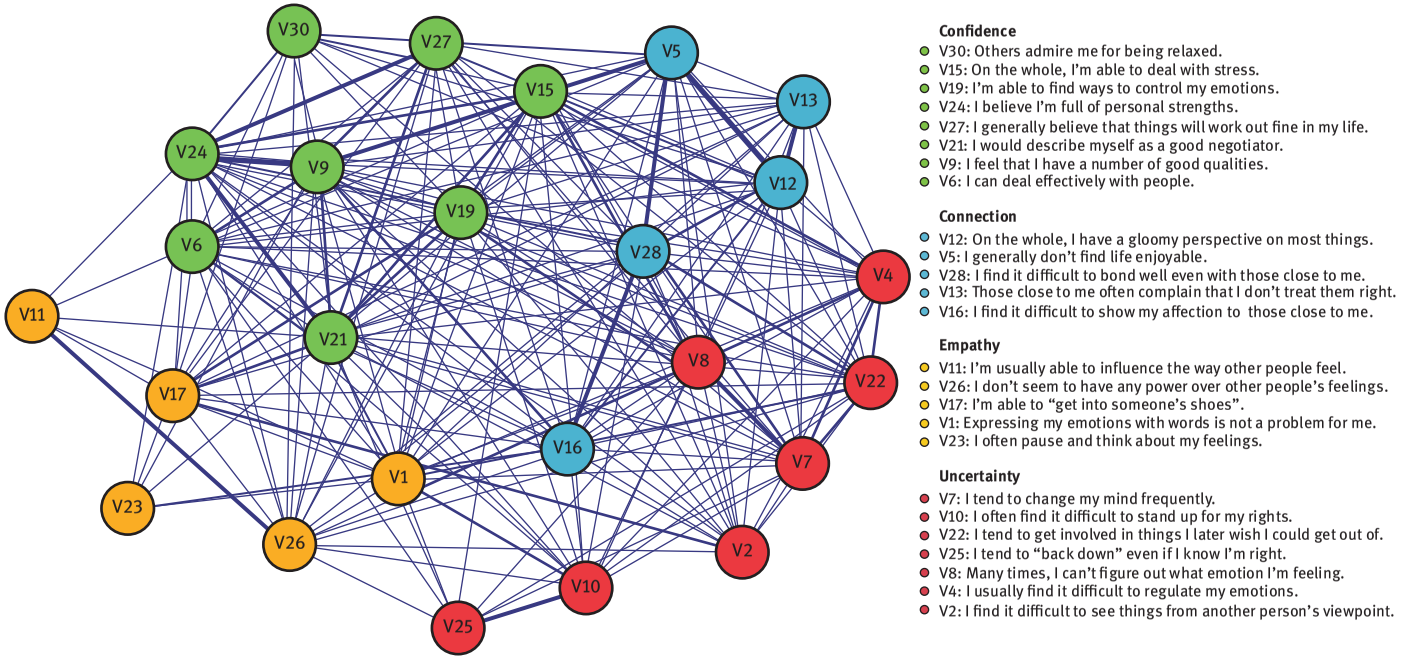

Fig. 15.1: Exploratory analysis of data using correlation graphs.

15.3.3 Step 3: building CFA model

We specified the laavan specific model, where each line represents a single latent factor with its indicators following the = ~ as shown in Box 15.4. In the code, we defined four latent factors referring to trait EI: confidence, connection, uncertainty, and empathy. We are assuming that 25 variables are indicators of those latent factors.

15.3.4 Step 4: examine model fit statistics

In the next step, a model was fitted to the TEIQue-SF data using the command model.fit from the laavan package, and is followed by printing a summary of the CFA results after fitting the model to the data as shown in Box 15.5. The laavan command summary provides a very extensive list of the CFA results with many details. However, most of the users will be interested in some of the basic CFA measures such as CFI, TLI, or RMSEA that can be obtained by a simple command fitMeasures as demonstrated in the supplementary materials in this chapter.

Box 15.4: CFA analysis using laavan package.

```
model <- 'Confidence = ~ V30 + V15 + V19 + V24 + V27 + V21 + V9 + V6
Connection = ~ V12 + V5 + V28 + V13 + V16
Uncertainty = ~ V7 + V10 + V22 + V25 + V8 + V4 + V2
Empathy = ~ V11 + V26 + V17 + V1 + V23'
```

Box 15.5: Examination of model fit, including visualization of the model.

```
model.fit <- cfa(model, teiq, std.lv = TRUE, missing = "fiml")
# Print model summary
summary(model.fit, fit.measures = TRUE, standardized = TRUE)
# Print only selected fit measures
fitMeasures(model.fit, c("cfi","tli", "rmsea"))
# Visualize normalized values for all items and corresponding factors
semPaths(model.fit, what = "std", layout = 'circle2', intercepts = FALSE, residuals = FALSE, nCharNodes = 10)
```

The fit of our four-factor model resulted in RMSEA = 0.058, CFI = 0.818 (NB: these values are identical to those obtained by Snowden et al. [16]), TLI = 0.797, and SRMR = 0.054 (NB: these values were not reported by Snowden et al.). Although some of the results like CFI or TLI are below the generally accepted threshold of 0.9, point at weak fit, one should be careful when relying on the CFA threshold values. As a final step of the CFA, we used semPaths command from the semPath package in R which can be used to visualize normalized values for all items and corresponding four factors (Fig. 15.2).

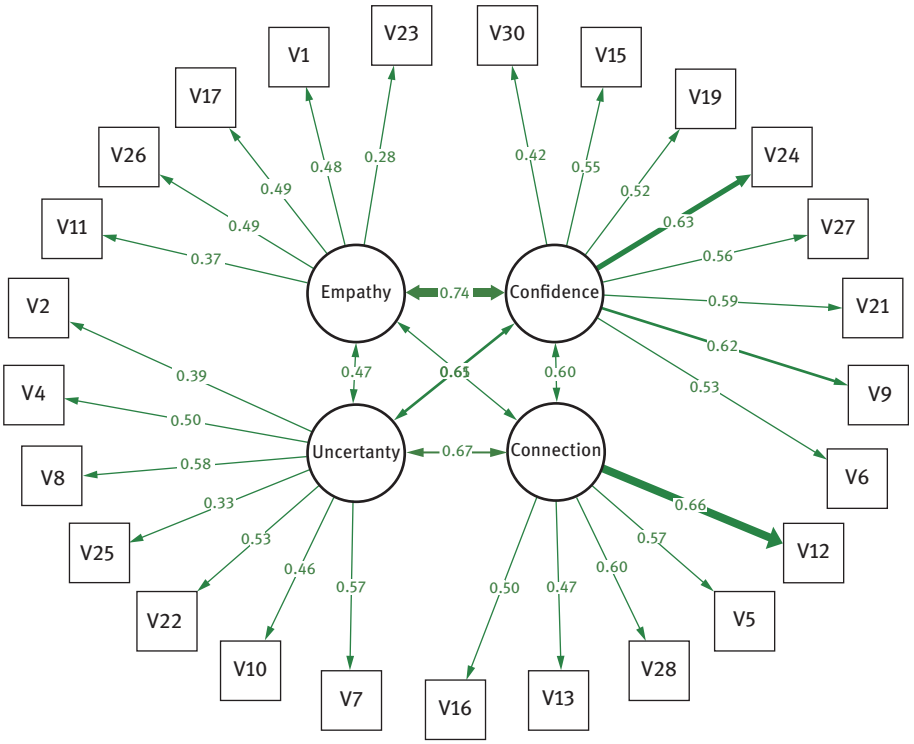

Fig. 15.2: Visualization of the obtained CFA model.

15.4 Discussion

The aim of research in nursing is the delivery of evidence-based care [37]. To collect and analyse the data, researchers need to have sufficient statistics, informatics and programming knowledge, and work experience. There are several software packages available for conducting various statistical analyses, which differ in capabilities for handling single group, multiple groups, non-normal variables, and missing data [38]. Although R is free to use, it is complex to use for clinicians with no statistical educational background. Thus, step-by-step tutorials and available scripts can help new or potential users in the transition from alternative, usually more costly, solutions.

Thus, we set out in this chapter to demonstrate, generally, the use of the R programming language for statistical analysis and, specifically, to demonstrate how to run a CFA using R [39]. Therefore, we were less concerned with establishing the structure of the TEIQue-SF, which had already been established in a previous study, than with demonstrating the utility of the R programming language. In this

light, the study was conducted against a background of very few "user-friendly" articles explaining how to use R for a specific analysis and especially to nurses who are not, traditionally, comfortable with the use of statistics [40]. Typical of the genre of articles purporting to introduce the use of R for a specific analysis is a study by Ritz and Striebig [41] in the *Journal of Statistical Software* which assumes considerable prior knowledge of programming generally and R specifically. Books and manuals on using R notoriously suffer from the same problems.

To conduct this study, we used a dataset that had already been analysed by CFA using the commercial package and IBM SPSS "bolt on" IBM programme AMOS (Analysis of Moment Structures) and published by Snowden et al. [16]. The differences between these two packages are striking. AMOS is modestly priced at under USD 100 but requires to be run alongside SPSS for which licenses cost more than USD 1,000. It should be noted that in some cases the price can be lower, especially with a student or campus-wide licensing. On the other hand, R is free to use. However, the facilities in AMOS are relatively much more user-friendly requiring only a minimum of technical skill to run [42]. The most attractive feature of AMOS is the ability to draw the path diagrams and, thereby, to visualize the structural equation models that are being analysed [38]. These can also be easily modified and the outputs – both graphical and numerical – are quite easy to understand, provided the user has a reasonable understanding of SEM, generally, and CFA, specifically. On the other hand, while we demonstrate that precisely the same outcomes for a CFA are obtained using either AMOS or R, the process of using R is considerably less intuitive. Unlike a commercial package such as AMOS – where the output includes all parameters including fit indices and a visual representation of the data – R requires multiple steps and the coordinated use of several different statistical packages. But R does have some additional features over AMOS and that the output of SEM can be easily visualized and in a form that is easily used in presentations and publications. Moreover, AMOS does not implement complex sampling estimation [43].

15.4.1 Limitations

We readily admit that R is not easy to master, and it is also customary for users only to make use of and have expertise in a limited range of the packages and methods available in R. Towards the end of helping other users, we provide in detail and in its entirety the R coding used in the present analysis, demonstrating how to load packages and subsequently to run specific aspects of the analysis. For others wishing to conduct CFA in R, we encourage simply by copying and pasting this coding into R or R Studio and we also make our database available for others to replicate this analysis. By publishing the code and data we also demonstrate that compared with most other alternatives, R offers a high level of reproducibility.

15.5 Conclusion

CFA is a common data-analytic method used in development of new measures and psychometric evaluation in nursing research. Thus, it is important that researchers have basic knowledge for conducting CFA analysis. CFA can be performed in R following basic steps as proposed in this chapter: reading and checking the data, conducting exploratory data analysis, building CFA model, and examining model fit statistics. R is an alternative statistical tool for conducting statistical analyses in nursing. R also ensures reproducible code and data, provides reproducibility of the results, and offers wide variety of visualization options.

References

[1] R Development Core Team. A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing, 2005.
[2] Beaujean AA. Factor analysis using R. Practical Assess Res Eval, 2013, 18(1), 4.
[3] Venables WN, Smith DM. R development core team. An introduction to R. Vienna, Austria, R Foundation for Statistical Computing, 2009.
[4] Ozgur C, Kleckner M, Li Y. Selection of statistical software for solving big data problems: A guide for businesses, students, and universities. SAGE Open, 2015, 5(2), 2158244015584379.
[5] Stiglic G, Watson R, Cilar L. R you ready? Using the R programme for statistical analysis and graphics. Res Nurs Health, 2019, 42(6), 494–499.
[6] Saha E, Ray PK. Statistical analysis of medical data for inventory management in a healthcare system. In: Analytics, operations, and strategic decision making in the public sector. IGI Global, 2019, 166–186.
[7] Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinform, 2008, 9(1), 559.
[8] Jones PJ, Mair P, McNally RJ. Visualizing psychological networks: A tutorial in R. Front Psychol, 2018, 9, 1742.
[9] Dima AL. Scale validation in applied health research: Tutorial for a 6-step R-based psychometrics protocol. Health Psychol Behav Med, 2018, 6(1), 136–161.
[10] Rietveld T, Van Hout R. Statistical techniques for the study of language behaviour. Berlin, Mouton de Gruyter, 1993.
[11] Matsunaga M. How to factor-analyze your data right: Do's, don'ts, and how-to's. Int J Psychol Stud, 2010, 3(1), 97–110.
[12] Brown TA. Confirmatory factor analysis for applied research. 2nd ed. New York, The Guildford Press, 2015.
[13] Child D. The essentials of factor analysis. 2nd ed. London, Cassel Educational Limited, 1990.
[14] Suhr DD Exploratory or confirmatory factor analysis? Proceedings of the 31st Annual SAS? Users Group International Conference. Cary, NC, SAS Institute. 2006.
[15] DeCoster J. Overview of factor analysis, 1998. (Accessed August 10, 2021, at: http://www.stat-help.com/notes.html ().
[16] Snowden A, Stenhouse R, Young J, Carver H, Carver F, Brown N. The relationship between emotional intelligence, previous caring experience and mindfulness in student nurses and midwives: A cross sectional analysis. Nurse Educ Today, 2015, 35(1), 152–158.

[17] Petrides KV. Psychometric properties of the Trait Emotional Intelligence Questionnaire. In: Stough C, Saklofske DH, Parker JD, eds. Advances in the assessment of emotional intelligence. New York, Springer, 2009.
[18] Freudenthaler HH, Neubauer AC, Gabler P, Scherl WG. Testing the Trait Emotional Intelligence Questionnaire (TEIQue) in a German-speaking sample. Pers Individ Differ, 2008, 45, 673–678.
[19] Petrides KV, Furnham A. The role of trait emotional intelligence in a gender-specific model of organizational variables. J Appl Soc Psychol, 2006, 36, 552–569.
[20] Zampetakis LA. Chapter 11: The measurement of trait emotional intelligence with TEIQue-SF: An analysis based on unfolding item response theory models'. In: Härtel CEJ, Ashkanasy NM, Zerbe WJ, eds. Research on emotion in organizations. 2011, vol. 7, 289–315.
[21] Stiglic G, Cilar Budler L, Watson R. Data for: Running a Confirmatory Factor Analysis in R: a step-by-step tutorial, Mendeley Data, V1, 2022; doi: 10.17632/bkh8wtgmkg.1.
[22] Albright JJ. Confirmatory factor analysis using AMOS, LISREL, and MPLUS. The Trustees of Indiana University, 2006.
[23] Salkind NJ. Encyclopedia of research design. vol. 1, Sage, 2010.
[24] Kyriazos TA. Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. Psychology, 2018, 9(08), 2207.
[25] Kaiser HF. A second generation little jiffy. Psychometrika, 1970, 35(4), 401–415.
[26] Brown TA, Moore MT. Confirmatory factor analysis. Handb Struct Equation Model, 2012, 361–379.
[27] Korkmaz S, Goksuluk D, Zararsiz G, An MVN. R package for assessing multivariate normality. R J, 2014, 6(2), 151–162.
[28] Finney SJ, DiStefano C. Non-normal and categorical data in structural equation modeling. In: Hancock GR, Mueller RO, eds. Structural equation modeling: A second course. 2nd ed. Charlotte, NC, Information Age Publishing, 2013, 439–492.
[29] Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network visualizations of relationships in psychometric data. J Stat Softw, 2012, 48(4), 1–18.
[30] JFJr H, Black WC, Babin BJ, Anderson RE, Tatham RL. Multivariate data analysis. 7th ed. Prentice-Hall, Upper Saddle River, 2010.
[31] Jackson DL, Gillaspy Jr JA, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: An overview and some recommendations. Psychol Methods, 2009, 14(1), 6.
[32] Tabachnick BG, Fidell LS. Using multivariate statistics. 5th ed. Boston, Allyn and Bacon, 2007.
[33] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equation Model, 1999, 6, 1–55.
[34] Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 1973, 38, 1–10.
[35] Chavez A, Koutentakis D, Liang Y, Tripathy S, Yun J. Identify Statistical Similarities and Differences Between the Deadliest Cancer Types Through Gene Expression. 2019.
[36] Shi D, Maydeu-Olivares A, Rosseel Y. Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. Struct Equ Model, 2020, 27(1), 1–15.
[37] Grove SK, Gray JR. Understanding nursing research E-book: Building an evidence-based practice. Elsevier Health Sci, 2018.
[38] Narayanan A. A review of eight software packages for structural equation modeling. Am Stat, 2012, 66(2), 129–138.
[39] R Core Team. A Language and Environment for Statistical Computing, 2013. (Accessed August 10, 2021), at: http://www.R-project.org.

[40] Hagen B, Awasoga O, Kellett P, Die O. Evaluation of undergraduate nursing students' attitudes towards statistics courses, before and after a course in applied statistics. Nurse Educ Today, 2013, 33, 949–955.
[41] Ritz C, Streibig JC. Bioassay analysis using R. J Stat Softw, 2005, 125.
[42] Arbuckle JL. IBM SPSS AMOS 20 user's guide. Armonk, IBM Corporation, 2011.
[43] Oberski D. lavaan. survey: An R package for complex survey analysis of structural equation models. J Stat Softw, 2014, 57(1), 1–27.