SPRINGER BRIEFS IN STATISTICS

Mark Stemmler

Person-Centered Methods

Configural Frequency Analysis (CFA) and Other Methods for the Analysis of Contingency Tables

Second Edition

Springer

SpringerBriefs in Statistics

More information about this series at http://www.springer.com/series/8921

Mark Stemmler

Person-Centered Methods

Configural Frequency Analysis (CFA) and Other Methods for the Analysis of Contingency Tables

Second Edition

Springer

Mark Stemmler
Institute of Psychology
Friedrich-Alexander University
of Erlangen-Nuremberg (FAU)
Erlangen, Germany

ISSN 2191-544X ISSN 2191-5458 (electronic)
SpringerBriefs in Statistics
ISBN 978-3-030-49420-9 ISBN 978-3-030-49421-6 (eBook)
https://doi.org/10.1007/978-3-030-49421-6

Preface to the Second Edition

The first edition was quite successful. I personally kept on teaching my person-centered method class at the College of Health and Human Development at the Pennsylvania State University (again many thanks to Linda Collins for her great hospitality in all those years. Linda is now leaving Penn State for New York University). I am still hosting CFA workshops not only in Germany (e.g., Erlangen, Cologne) and the US (e.g., Temple University, Lafayette College) but also in China (e.g., Nanjing and Chongqing). Those workshops keep me close to the questions and obstacles faced by the users of CFA.

After six years, this second edition was necessary. Next to correcting typos and other errors of the first edition, CFA and the wonderful R-package written by Dr. Jörg-Henrik Heine, who is now at the Technical University Munich (TUM), have developed further. The R package *confreq* 1.5.5 was written and adjusted to the latest R version 4.0. In addition, many useful things can now be analyzed with the R-package (e.g., Configural Mediation Models, two-sample CFA, nonstandard CFA). A new α-protection was implemented (i.e., the Holmes protection) and the design matrix now can be easily extended or modified. I am sure that this second edition will find more and more recognition in the community of researchers using the person-centered approach many thanks to Jörg! I also say thank you to Alexander von Eye, who was always very helpful in answering any upcoming questions regarding CFA. Amanda Applegate was again very helpful for proofreading my Germanic English, thanks a million. My deep thanks go out to Laura Briskman from the Springer Publishing Company in New York for facilitating the publication of this second edition.

Finally, I offer my deepest thanks to my wife Susanne. Fate wasn't always nice to us in the past few years, but we both managed to cope with those extremely difficult times. I also thank my beloved son Quincy, who is now happily married to Cristina Colmenares and who lives in Lima (Peru). My whole family gives me comfort and a secure base, but they also provide me with energy for new scientific projects like this second edition.

This book is published together with an electronic supplement containing 18 R-files (syntax and data), which are described in the book and which can be run in order to become an expert in running the R-package *confreq*.

Erlangen, Germany Mark Stemmler
Spring 2020

Preface to the First Edition

The idea for this book came to me while I was teaching courses during the summer at The Methodology Center at Penn State (Director: Linda Collins). Teaching classes on person-centered methods which do not belong to the standard curriculum at German or American universities was very inspiring to me. The interaction with the students helped me to understand how to explain the content of this class so that it is easy to understand and showed to me how much the students liked this different look at statistics.

This book will take an easy-to-understand look at the statistical approach called the *person-centered method*. Instead of analyzing means, variances and covariances of scale scores as in the common variable-centered approach, the person-centered approach analyzes persons or objects grouped according to their characteristic patterns or configurations in contingency tables. The main focus of the book will be on *Configural Frequency Analysis* (CFA; Lienert and Krauth, 1975). In complex contingency tables, patterns or configurations are analyzed while comparing observed cell frequencies with expected frequencies. Significant differences between observed and expected frequencies lead to the emergence of *types* and *antitypes*. Types are patterns or configurations which are significantly more often observed than the expected frequencies; antitypes represent configurations which are observed less frequently than expected. CFA is very much related to log-linear modeling. In log-linear modeling the goal is to come up with a fitting model including all important variables. Instead of fitting a model, CFA looks at the significant residuals of a log-linear model.

CFA was invented by Gustav A. Lienert, an Austrian physician and professor of psychology, who died in 2001. I was lucky to have met Gustav A. Lienert, who was a very inspiring and enthusiastic person. I am thankful for his cheerfulness and his support. I was introduced to 'Herrn Lienert' by Alexander von Eye (Psychology Professor at Michigan State and University of Vienna). I am very thankful to Alex who has introduced me to the field of categorical data analysis.

A number of ideas presented here (especially those in Chapter 6) were proposed by Erwin Lautsch. They were all published in a series of Special Issues on CFA (guest editor together with Alexander von Eye) in the German journal called

Psychology Science (formerly known as the *Psychologische Beiträge*). Thank you Erwin for sharing your ideas!

One important asset to this book was the development of the R-package *confreq* (derived from **con**figural **freq**uency analysis). The open source software R is available at no cost and is developing in a fast and progressive manner. An R-package was also important because there was no readily available software for configural frequency analysis (with exception of a somewhat outdated DOS software written in FORTRAN). *Confreq* was written by Jörg-Henrik Heine (LMU Munich). I met Jörg at our annual statistical meetings in Rothenberge (Northern Germany) organized by Christian Tarnai and Jost Reinecke. Jörg worked diligently on this package for more than two years including several setbacks. Many thanks to you Jörg! I am also thankful to Rainer Alexandrowicz (who I also met in Rothenberge) who worked on Stirlings's formula for using the binomial test as part of *confreq*.

My thanks go out to Amanda Applegate and Heather Foran for proofreading my Germanic English. In addition, Heather also addressed to me all the relevant sections which were difficult to understand and not well explained. Her methodological perspective was extremely essential for my writing! Thank you to Hannah Bracken at Springer for her support in leading my book endeavor.

Finally, I offer my deepest thanks to my wife Susanne and my son Quincy. Thanks for giving me so much comfort and for energizing my life.

Erlangen, Germany Mark Stemmler
Spring 2014

Contents

Chapter 1
Introducing Person-Centered Methods

Abstract This chapter explains the term *person-centered methods* and how ***configural frequency analysis*** **(CFA)** works. Instead of analyzing means, variances, and covariances of scale scores as in the common variable-centered approach, the person-centered approach analyzes persons or objects grouped according to their characteristic configurations in complex contingency tables while comparing observed cell frequencies with expected frequencies. CFA is a statistical method that looks for over- and under-frequented cells. Over-frequented indicates that the observations in this cell or configuration are observed more often than expected, and under-frequented indicates that this configuration is observed less often than expected. In CFA, a pattern or configuration that contains significantly more observed cases than expected is called a *type*; similarly, a configuration that contains significantly fewer observed cases than expected is called an *antitype*. In addition, this chapter includes an explanation of Meehl's paradox [12], which postulates that it is possible to have a bivariate relationship with a zero association or correlation and, at the same time, a higher order association or correlation. Meehl argued for investigating higher order interactions (beyond bivariate interactions), which can be detected with CFA.

1.1 What is Configural Frequency Analysis (CFA) Good for?

This chapter takes an accessible look at the statistical approach called the *person-centered methods*. The person-centered approach analyzes units of interest (e.g., persons, animals, or objects) grouped according to their characteristic patterns or configurations in cross-tabulations (see [1, 2, 16, 19]). The observed patterns are arranged in tables called contingency tables, ordered by their indices. A certain position in such a table, denoted by a pattern or configuration, is called a cell [21]. The main focus of this book will be on ***Configural Frequency Analysis*** (CFA; [9, 18, 23, 24]), which is a statistical method that compares observed cell frequencies with expected frequencies while looking for over- and under-frequented cells or configurations. Over-frequented indicates that the observations in this cell or configuration

M. Stemmler, *Person-Centered Methods*,
SpringerBriefs in Statistics,
https://doi.org/10.1007/978-3-030-49421-6_1

are observed more often than expected, and under-frequented indicates that this configuration is observed less often than expected. In CFA, a configuration that contains more observed cases than expected is called a *type*; similarly, a configuration that contains fewer observed cases than expected is called an *antitype*. CFA was invented by Gustav A. Lienert, an Austrian physician and professor of psychology, who died in 2001 [11, 20]. CFA is similar to log-linear modeling. In log-linear modeling, the goal is to come up with a fitting model including all important variables. Instead of fitting a model, CFA examines the significant residuals of a log-linear model.

1.2 What are Typical Research Questions That can be Answered by CFA?

To orient the reader to think in terms of patterns or configurations, we begin with several research examples based on the use of CFA. The examples also demonstrate the frequent interdisciplinary usage of CFA:

(1) Take an example from hydrobiology [13]. Let's say a researcher is interested in fish habitats or, specifically, spawning habitats of fish, because a sufficient fish stock is important for the ecological system of a river. In logistic regression, the researcher compares places with many fish with places with a few fish. Based on logistic regression or log-linear modeling, researchers know different important features of the river such as flow velocity, type of structure and substrate of the river bed, and the vegetation of the riverbanks, but they don't know the optimal combination of the features resulting in a typical (i.e., over-frequented) fish habitat.[1] With CFA, one can identify significant cell configurations, providing CFA answers at the level of individual cells (configurations) instead of at the level of variables.

(2) Another research question is from the field of pediatrics. In a small sample of premature newborns with additional neurological or other health problems (e.g., seizures, need for mechanical ventilation) a researcher uses CFA to search for characteristic configurations that predict normal (i.e., typical) cognitive development at age 5.[2]

(3) In an example taken from *veterinary medicine* [6], Turkish researchers used CFA to investigate the causes of the death of cattle calves. They recorded the characteristics of the barn system (separation of mothers and calves or joint rearing), type of disease (intestinal disease, respiratory disease, trauma), vaccination status (vaccinated versus unvaccinated), and sex. They found a significant configuration: calves die more often than expected if they have an intestinal disease,

[1] By the way, many European fish like a shaded habitat with a fine and coarse substrate, depending on high flow velocity.

[2] Prematurely born babies who were mechanically ventilated but who had no seizures have the best chances for normal cognitive development.