

The Cressie-Read Family of Divergence Statistics and Correspondence Analysis

Authors:

Rosaria Lombardo – University of Campania (Italy) [presenting]

Eric Beh – University of Wollongong (Australia), Stellenbosch University (South Africa)

Keywords:

Cressie-Read family of divergence statistics, correspondence analysis, data visualisation

Abstract:

Correspondence analysis is a popular technique for visualising the association between categorical variables and has been extensively studied (Greenacre, 1984; Beh and Lombardo, 2014). It generally involves using Pearson's chi-squared statistic as its numerical foundation, however, the statistic is only one member of the Cressie-Read family of divergence statistics (Cressie and Read, 1984) which also includes the Freeman-Tukey statistic, the likelihood ratio statistic, and modified versions of these measures. Therefore, we discuss a general framework that explores how correspondence analysis can be performed using this family to suite different data structures (Beh and Lombardo, 2024). Special cases of the framework include the correspondence analysis using the Freeman-Tukey statistic (Beh et al., 2018) which is linked to the Hellinger distance decomposition (Cuadras and Cuadras, 2006) method, and the log-likelihood ratio statistic, which leads to log-ratio analysis (Greenacre, 2010).

Additionally, we will explore some modifications to this framework, including its application to multi-way correspondence analysis (Lombardo et al., 2021).

References

- Beh, E.J. and Lombardo, R. (2014). *Correspondence analysis: Theory, Practice and New Strategies*. Wiley, Chichester.
- Beh, E. J., Lombardo, R. and Alberti, G. (2018). Correspondence analysis and the Freeman-Tukey statistic: A study of archaeological data. *Computational Statistics and Data Analysis*, 128, 73 – 86.
- Beh E. J., Lombardo R. (2024) Correspondence Analysis Using the Cressie–Read Family of Divergence Statistics. *International Statistical Review*, 91,1, 17-42 doi: 10.1111/insr.12541
- Cressie, N. A. C. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, 46, 440 – 464.
- Cuadras, C. M. and Cuadras D. (2006). A parametric approach to correspondence analysis. *Linear Algebra and its Applications*, 417, 64 – 74.
- Greenacre, M. J. (1984). *Theory and Application of Correspondence Analysis*. Academic Press, London.
- Greenacre, M. (2010). Log-ratio analysis is a limiting case of correspondence analysis. *Mathematical Geosciences*, 42, 129 – 134.
- Lombardo, R., Beh, E.J. and Kroonenberg P.M. (2021). Symmetrical and non-symmetrical variants of three-way correspondence analysis for ordered variables. *Statistical Science*, 36, 542 – 561.