

Scealy, J. L.; Welsh, A. H.

Colours and cocktails: compositional data analysis 2013 Lancaster lecture. (English)

Zbl 1336.62028

Aust. N. Z. J. Stat. 56, No. 2, 145-169 (2014).

Summary: The different constituents of physical mixtures such as coloured paint, cocktails, geological and other samples can be represented by d -dimensional vectors called compositions with non-negative components that sum to one. Data in which the observations are compositions are called compositional data. There are a number of different ways of thinking about and consequently analysing compositional data. The log-ratio methods proposed by Aitchison in the 1980s have become the dominant methods in the field. One reason for this is the development of normative arguments converting the properties of log-ratio methods to 'essential requirements' or Principles for any method of analysis to satisfy. We discuss different ways of thinking about compositional data and interpret the development of the Principles in terms of these different viewpoints. We illustrate the properties on which the Principles are based, focussing particularly on the key subcompositional coherence property. We show that this Principle is based on implicit assumptions and beliefs that do not always hold. Moreover, it is applied selectively because it is not actually satisfied by the log-ratio methods it is intended to justify. This implies that a more open statistical approach to compositional data analysis should be adopted.

MSC:

62-07 Data analysis (statistics) (MSC2010)

62-02 Research exposition (monographs, survey articles) pertaining to statistics

Cited in 5 Documents

Keywords:

amalgamation; log-ratio transformation; principles of compositional data analysis; subcomposition; subcompositional coherence; square root transformation

Full Text: DOI

References:

- [1] Aitchison, The statistical analysis of compositional data (with discussion), *J. R. Stat. Soc. B* 44 pp 139– (1982) · Zbl 0491.62017
- [2] Aitchison, The statistical analysis of geochemical compositions, *J. Math. Geol.* 16 pp 531– (1984) · doi:10.1007/BF01029316
- [3] Aitchison, The Statistical Analysis of Compositional Data (1986) · Zbl 0688.62004 · doi:10.1007/978-94-009-4109-0
- [4] Aitchison, Reply to 'Interpreting and testing compositional data' by Alex Woronov, Karen M. Love and John C Butler, *J. Math. Geol.* 21 pp 65– (1989) · doi:10.1007/BF00897241
- [5] Aitchison, Letter to the Editor. Comment on 'Measures of variability for geological data' by D.F. Watson and G.M. Philip, *J. Math. Geol.* 22 pp 223– (1990) · doi:10.1007/BF00891826
- [6] Aitchison, Letter to the Editor. Delusions of uniqueness and ineluctability, *J. Math. Geol.* 23 pp 275– (1991) · doi:10.1007/BF02066299
- [7] Aitchison, On criteria for measures of compositional distance, *J. Math. Geol.* 24 pp 365– (1992) · Zbl 0970.86531 · doi:10.1007/BF00891269
- [8] Aitchison, J. 2003 A concise guide to compositional data analysis
- [9] Aitchison, Reply to Letter to the Editor by S. Rehder and U. Zier on 'Logratio analysis and compositional distance' by J. Aitchison, C. Barceló-Vidal, J.A. Martín-Fernández and V. Pawlowsky-Glahn, *J. Math. Geol.* 33 pp 849– (2001) · Zbl 1101.86310 · doi:10.1023/A:1010954915624
- [10] Aitchison, Logratio analysis and compositional distance, *J. Math. Geol.* 32 pp 271– (2000) · Zbl 1101.86309 · doi:10.1023/A:1007529726302
- [11] Aitchison, Some comments on compositional data analysis in archaeometry, in particular the fallacies in Tangri and Wright's dismissal of logratio analysis, *Archaeometry* 44 pp 295– (2002) · doi:10.1111/1475-4754.t01-1-00061
- [12] Aitchison, Compositional data analysis: Where are we and where should we be heading?, *J. Math. Geol.* 37 pp 829– (2005) · Zbl 1177.86017 · doi:10.1007/s11004-005-7383-7
- [13] Aitchison, Logistic-normal distributions: Some properties and uses, *Biometrika* 67 pp 261– (1980) · Zbl 0433.62012 · doi:10.2307/2335470
- [14] Atkinson, *J. R. Stat. Soc. B* 44 pp 139– (1982)

- [15] Bacon-Shone, *Compositional Data Analysis: Theory and Applications* pp 3– (2011)
- [16] Butler, A latent Gaussian model for compositional data with zeros, *Appl. Statist.* 57 pp 505– (2008)
- [17] Chayes, On ratio correlation in petrology, *J. Geol.* 57 pp 239– (1949) · doi:10.1086/625606
- [18] Chayes, On correlation between variables of constant sum, *J. Geophys. Res.* 65 pp 4185– (1960) · doi:10.1029/JZ065i012p04185
- [19] Egozcue, Reply to 'On the Harker variation diagrams;...' by J.A. Cortés, *J. Math. Geol.* 41 pp 829– (2009)
- [20] Egozcue, *Compositional Data Analysis: Theory and Applications* pp 12– (2011) · doi:10.1002/9781119976462.ch2
- [21] Filzmoser, *Compositional Data Analysis: Theory and Applications* pp 59– (2011) · doi:10.1002/9781119976462.ch5
- [22] Greenacre, Power transformations in correspondence analysis, *Comp. Stat. Data Anal.* 53 pp 3107– (2009) · Zbl 1453.62099 · doi:10.1016/j.csda.2008.09.001
- [23] Greenacre, *Compositional Data Analysis: Theory and Applications* pp 12– (2011)
- [24] Greenacre, Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio-scale measurements, *J. Classification* 6 pp 29– (2009) · Zbl 1276.62037 · doi:10.1007/s00357-009-9027-y
- [25] Grunsky, E. 2012 Using multi-element geochemical data for process discovery and geologic mapping - From mess to message
- [26] Lancaster, The Helmert matrices, *Am. Math. Monthly* 72 pp 4– (1965) · Zbl 0124.01102 · doi:10.2307/2312989
- [27] Leininger, Spatial regression modeling for compositional data with many zeros, *J. Agric. Biol. Environ. Stat.* 18 pp 314– (2013) · Zbl 1303.62085 · doi:10.1007/s13253-013-0145-y
- [28] Martín-Fernández, *Proceedings of IAMG '98 - The Fourth Annual Conference of the International Association of Mathematical Geology* pp 526– (1998)
- [29] Martín-Fernández, *Compositional Data Analysis: Theory and Applications* pp 43– (2011) · doi:10.1002/9781119976462.ch4
- [30] Pearson, Mathematical contributions to the theory of evolution - On a form of spurious correlation which may arise when indices are used in the measurement of organs, *Proc. R. Soc. Lon.* 60 pp 489– (1897) · Zbl 28.0209.02 · doi:10.1098/rspl.1896.0076
- [31] Rayens, Box-Cox transformations in the analysis of compositional data, *J. Chemometrics* 5 pp 227– (1991a) · doi:10.1002/cem.1180050310
- [32] Rayens, Estimation in compositional data analysis, *J. Chemometrics* 5 pp 361– (1991b) · doi:10.1002/cem.1180050405
- [33] Rehder, Letter to the Editor. Comment on 'Logratio analysis and compositional distance' by Aitchison et al. (2000), *J. Math. Geol.* 32 pp 845– (2001) · Zbl 1011.86504 · doi:10.1023/A:1010902931554
- [34] Scealy, *Proceedings of CoDaWork'11: 4th International Workshop on Compositional Data Analysis* pp 10– (2011a)
- [35] Scealy, Regression for compositional data by using distributions defined on the hypersphere, *J. R. Stat. Soc. B* 73 pp 351– (2011b) · doi:10.1111/j.1467-9868.2010.00766.x
- [36] Scealy, Fitting Kent models to compositional data with small concentration, *Stat. Comput.* (2012) · Zbl 1325.62049
- [37] Stanley, Descriptive statistics for N-dimensional closed arrays: a spherical co-ordinate approach, *J. Math. Geol.* 22 pp 993– (1990)
- [38] Stephens, Discussion of Aitchison, J. (1982). The statistical analysis of compositional data (with discussion), *J. R. Stat. Soc. B.* 44 pp 139– (1982a)
- [39] Stephens, Use of the von Mises distribution to analyse continuous proportions, *Biometrika* 69 pp 197– (1982b) · doi:10.1093/biomet/69.1.197
- [40] Stewart, Managing the essential zeros in quantitative fatty acid signature analysis, *J. Agric. Biol. Environ. Stat.* 16 pp 45– (2010) · Zbl 1306.62237 · doi:10.1007/s13253-010-0040-8
- [41] Thió-Henestrosa, S. Martín-Fernández, J. 2003 *Proceedings of CoDaWork'03, The 1st Compositional Data Analysis Workshop* <http://ima/udg.es/Activitats/CoDaWork03>
- [42] Tsagris, *Proceedings of CoDaWork'11: 4th International Workshop on Compositional Data Analysis* (2011)
- [43] Watson, Letter to the Editor. Reply to Comment on 'Measures of variability for geological data' by D.F. Watson and G.M. Philip, *J. Math. Geol.* 22 pp 227– (1990) · doi:10.1007/BF00891827
- [44] Watson, Letter to the Editor. Reply to "Delusions of uniqueness and ineluctability" by J. Aitchison, *J. Math. Geol.* 23 pp 279– (1991) · doi:10.1007/BF02066300
- [45] Watson, Measures of variability for geological data, *J. Math. Geol.* 21 pp 233– (1989) · doi:10.1007/BF00893217
- [46] Zier, *Proceedings of IAMG98, The Fourth Annual Conference of the International Association for Mathematical Geology* pp 555– (1998)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.