



Data Article

Data on Gaussian copula modelling of the views of sport club members relating to community sport, Australian sport policy and advocacy



Charles Mountifield*

Discipline of Sport and Exercise Science, Faculty of Health, University of Canberra, 11 Kirinari St, Bruce ACT 2617, Australia

ARTICLE INFO

Article history:

Received 25 December 2021

Revised 22 March 2022

Accepted 24 March 2022

Available online 28 March 2022

Dataset link: [GCGM Data \(Original data\)](#)*Keywords:*

Advocacy

clubs, community sport
policy

sustainability

ABSTRACT

In Australia, community sport plays an essential role in the objectives of national sports organizations and local and national social policy more generally. The opinions of senior community sport club officials ($n=53$) in rural New South Wales concerning sustainability, policy and advocacy matters impacting upon community sports clubs (CSCs) were identified. Participants were surveyed to establish the level of influence of top-down sport management structures alongside rudimentary views on a coalition to advocate for community sport. Gaussian copula graphical models (GCGMs) were applied to demonstrate the raw and partial correlations between participant responses. While frequently associated with finance, science and medicine, the application of copulas is increasingly common in research in sport where there are multiple variables and relationships involved. GCGMs have been used in, for example, analysis in soccer and across physical education in general.

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

* Corresponding author.

E-mail address: charles.mountifield@canberra.edu.au

Specifications Table

Subject	Public Health and Health Policy
Specific subject area	Community sport policy
Type of data	Figures, Table
How the data were acquired	Survey, using Qualtrics (copy attached)
Data format	Raw
Description of data collection	Data was collected through surveys (using Qualtrics) of community sport clubs.
Data source location	University of Canberra, Canberra, Australia
Data accessibility	Mendeley Data DOI: 10.17632/j7sg3z3tf4.3 https://data.mendeley.com/datasets/j7sg3z3tf4/3

Value of the Data

- Descriptive analytics are useful to understand patterns, trends, and anomalies in data.
- The software package and the example data demonstrate that the GCGM modeling is useful for the analysis of a broad range of data types.
- GCGMs can be used to perform research in sport policy where there is a significant difference in demographic variables including geographical regions, sports, and CSCs.
- The application of GCGMs is increasingly common in research in sport.
- Policy-makers can benefit from the data in the policy creation and implementation process.

1. Data Description

With a focus on the specific subject areas of resource, policy and advocacy, the data collected produced results representing: (i) Descriptive statistics (percentages and 95% confidence intervals) of responses for each statement; (ii) raw and partial correlations between each statement; (iii) the 'loadings' from the PCA for each theme; and (iv) raw and partial correlations between each principal component. The results are outlined in the following four figures and supporting descriptions:

1.1. Descriptive statistics (percentages and 95% confidence intervals) of responses for each statement

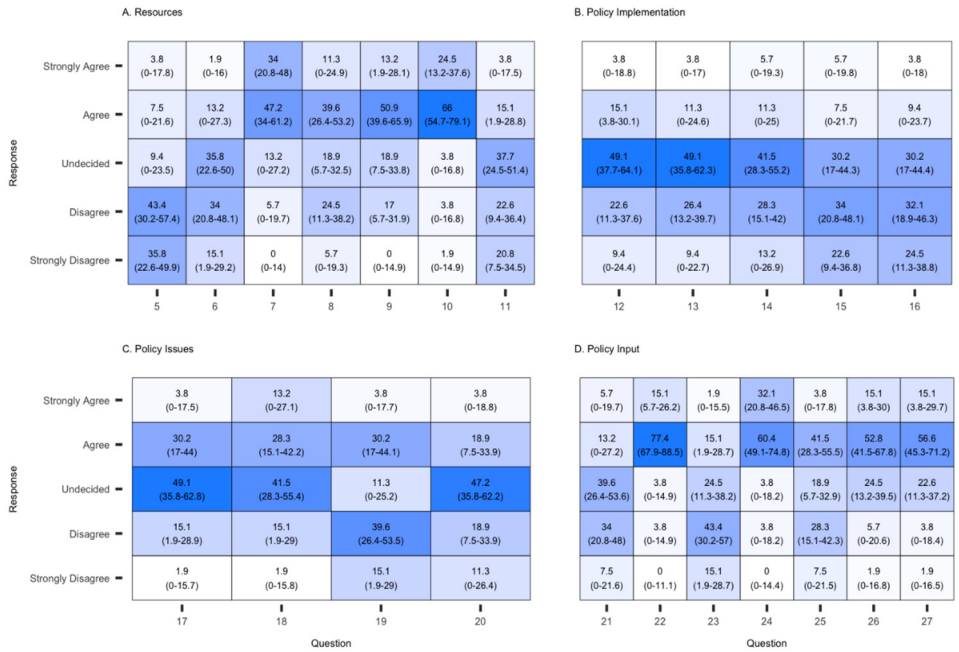


Fig. 1. Percentages and 95% confidence intervals (presented in parentheses) of responses for each statement (the darker shaded cells indicate a higher percentage of responses for a given statement).

Fig. 1 outlines descriptive statistics – the percentages and 95% confidence intervals – of responses for each statement of the four ‘themes’. A key initial observation is that 79.2% of CSC members surveyed indicated that their club had not been allocated adequate funding (Q5) and 90.5% agreed or strongly agreed that shared resources would enhance the relationships between clubs (Q10) (Fig. 1A). Only a minority of participants (13.2 – 18.9%) agreed or strongly agreed that NSO communication and coordination regarding policy implementation was effective (Fig. 1B). Several participants (30.2 – 49.1%), however, were undecided regarding this series of statements (Q12 – 16). 41.5% of senior CSC officials indicated that NSOs prioritized performance sport over participation (Q18). In comparison, 47.2% of participants were also undecided if their NSO offered clear channels for policy dispute resolution (Q20) (Fig. 1C). Notably, while 92.5% of participants agreed or strongly agreed that it is important to play an active role in shaping national sport-related policy (Q22), only 17% indicated that their club achieved this (Q23) (Fig. 1D).

1.2. Raw and partial correlations between each statement

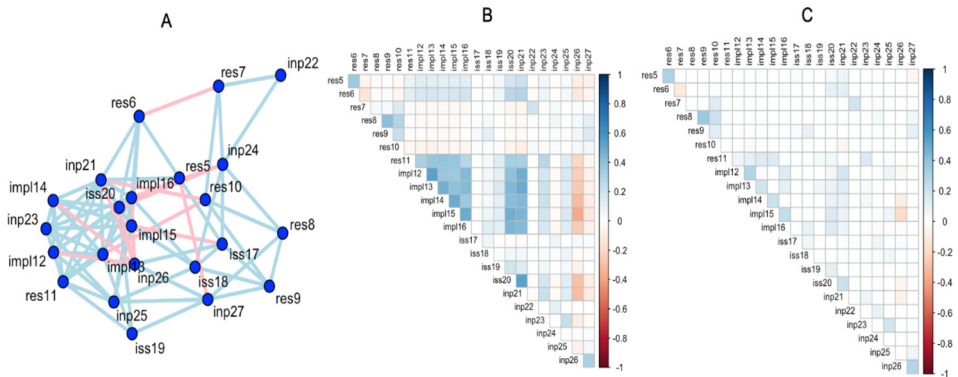


Fig. 2. Raw and partial correlations between each statement. The labels for the figures are abbreviated labels, with the first three letters indicating which 'theme' the statement belongs to, followed by the statement number. Fig. 3A highlights the direct relationships between statements with a **BLUE** line indicating a positive relationship and a **RED** line showing a negative relationship. Indirect relationships are present if two statements are joined by a third linking statement. For example, res8 and res9 are directly related to a positive relationship, whereas inp27 and res8 are indirectly related by each being directly associated with res9. Fig. 2B shows the raw correlations between statements. Fig. 2C presents the conditional relationships (partial correlations) between statements.

Fig. 2C reveals the following notable partial correlations between individual statements:

- i. A small positive relationship between res5 (Q.5 – Adequate funding has been allocated to my club) and res6 (Q.6 – My club has sufficient resources to achieve NSO/government policy objectives for community sport).
- ii. A small negative relationship between res6 and res7 (Q.7 – Access to improved management systems would benefit my club).
- iii. A moderate positive relationship between res8 (Q.8 – My club has resources it could share with other clubs) and res9 (Q.9 – Other clubs have resources they could share with my club).
- iv. A small to moderate positive relationship between impl12 (Q.12 – The NSO provides clear guidelines on how national plans will be implemented locally) and impl13 (Q.13 – The NSO provides a clear management structure for policy implementation at a local level).
- v. A small negative relationship between impl15 (Q.15 – NSOs, government and my club communicate regularly to ensure coordinated delivery of activities designed to grow participation in sport) and inp26 (Q.26 - My club would benefit from support from a local advocacy group that is independent of the NSO and government).

1.3. Principal components analysis (PCA)

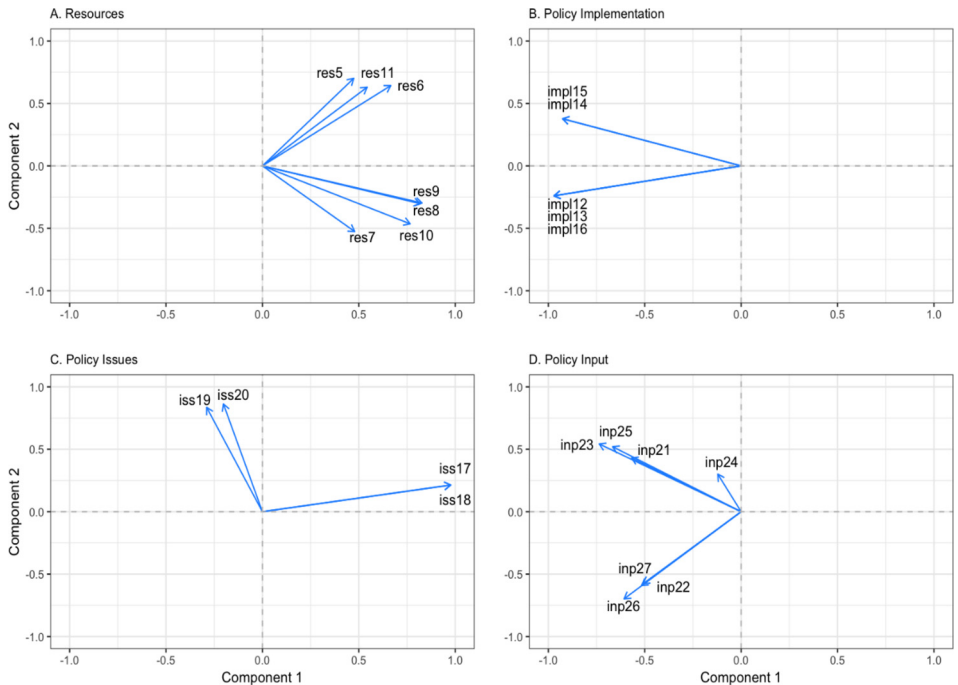


Fig. 3. Loadings from the principal components analysis (PCA) for each theme of statements. The labels for the figures are abbreviated labels, with the first three letters indicating which 'theme' the statement belongs to, followed by the statement number. The further an arrow moves along the x-axis towards 1, the more positively related that statement is to the first principal component (new artificial variable 1). The further an arrow moves along the x-axis towards -1, the more negatively related that statement is to the first principal component. The further an arrow moves along the y-axis towards 1, the more positively related that statement is to the second principal component. The further an arrow moves along the y-axis towards -1, the more negatively related that statement is to the second principal component. Thus, arrows/ statements that are more closely positioned to each other are more closely related.

Fig. 3 demonstrates the 'loadings' from the PCA for each theme. In terms of the PCA results relating to each theme, these are noted as follows:

- For Resources: two principal components (PCs) were formed, with the first PC (res1 in Fig. 4) explaining 44.8% of the variance, and the second PC (res2 in Fig. 4) explaining 28.2% of the variance, collectively explaining 73% of the variance of the original statements. The first PC was positively correlated with all statements regarding resources (Q.5 – Q.11), providing an artificial variable to describe a general agreement with each of these statements, or vice versa, a general disagreement with each of these statements. The second PC was positively correlated with Q.5 (Adequate funding has been allocated to my club), Q.6 (My club has sufficient resources to achieve NSO/government policy objectives for community sport) and Q.11 (My club uses a performance management system to ensure that it adheres to government and NSO sport policy), but negatively correlated with Q.7 (Access to improved management systems would benefit my club), Q.8 (My club has resources it could share with other clubs), Q.9 (Other clubs have resources they could share with my club), and Q.10 (Shared resources would enhance relationships between clubs). Notably, however, the correlations with Q.8 and Q.9 appear weak in strength (Fig. 3A).
- For Implementation: two PCs were formed, with the first PC (impl1 in Fig. 4) explaining 90.8% of the variance and the second PC (impl2 in Fig. 4) explaining 9.2% of the variance,

- collectively explaining 100% of the variance of the original statements. The first PC was negatively correlated with all statements regarding Implementation (Q.12-16) (Fig. 3B).
- iii. For Issues: two PCs were formed, with the first PC (iss1 in Fig. 4) explaining 50.8% of the variance and the second PC (iss2 in Fig. 4) explaining 38.4% of the variance, collectively explaining 89.1% of the variance of the original statements. The first PC is positively and strongly correlated with Q.17 (There are frequent differences between my club and the NSO regarding the priorities for local-level community sport) and Q.18 (The NSO prioritizes performance sport over programs to increase participation levels in sport). The second PC is positively and strongly correlated with Q.19 (Policy implementation is the sole responsibility of my club) and Q.20 (My NSO provides clear channels for policy dispute resolution).
 - iv. For Input: three PCs were formed, with the first PC (inp1 in Fig. 4) explaining 31.7% of the variance, the second PC (inp2 in Fig. 4) explaining 28.7% of the variance, and the third PC (inp3) explaining 16.7%, collectively explaining 77.2% of the variance of the original statements. The first PC was negatively correlated with all statements related to Input (Q.21-25). However, Q.24 (It is important for clubs to play an active role in shaping national sport-related policy) exhibited a weak relationship. The second PC was positively correlated with Q.21 (There are mechanisms in place for feedback and review between the NSO and my club), Q.23 (My club plays an active role in shaping national sport-related policy) and Q.25 (My club plays an active role in shaping local sport-related policy), and was negatively correlated with Q.22 (It is important for clubs to play an active role in shaping national sport-related policy), Q.26, and Q.27. The third PC was positively correlated with Q.22 and Q.24 and negatively correlated with Q.21 and Q.27.

1.4. Raw and partial correlations between each principal component

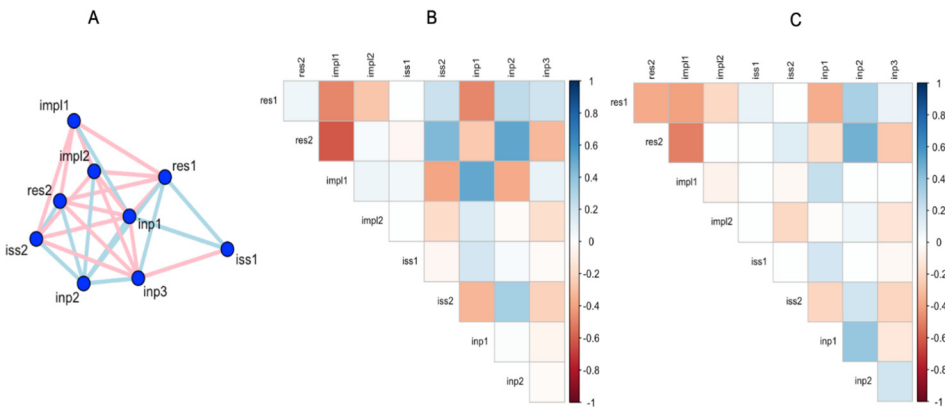


Fig. 4. Raw and partial correlations between each principal component. The labels for the figures are abbreviated labels, with the first three letters indicating which 'theme' the principal component belongs to, followed by the principal component number. Figure 4A highlights the direct relationships between principal components, with a **BLUE** line indicating a positive relationship and a **RED** line indicating a negative relationship. Indirect relationships are present if two statements are joined by a third linking statement. For example, inp2 and inp3 are directly related to a positive relationship, whereas iss1 and inp2 are indirectly related by each being directly related to inp3. Figure 4B shows the raw correlations between principal components. Fig. 4C presents the conditional relationships (partial correlations) between principal components. The labels for the figures are abbreviated labels, with the first three letters indicating which 'theme' the principal component belongs to, followed by the principal component number.

Fig. 4C revealed several notable partial correlations between principal components in different themes as described below:¹

- i. A moderately negative relationship between PC res1 and PC impl1, indicating that those senior CSC officials that generally disagreed with Q.5-11 (Resources) also tended to disagree with Q.12-16 (Implementation) and vice versa.
- ii. A moderately negative relationship between PC res2 and PC impl1, indicating that those senior CSC officials that disagreed with Q.5 (Adequate funding has been allocated to my club), Q.6 (My club has sufficient resources to achieve NSO/government policy objectives for community sport) and Q.11 (My club uses performance management system to ensure that it adheres to government and NSO sport policy), and agreed with Q.7-10 (regarding shared resources and access to improved management systems), tended to also disagree with Q.12-16 (regarding implementation) and vice versa.
- iii. A moderately negative relationship between PC res1 and PC inp1, indicating that those senior CSC officials that generally disagreed with Q.5-11 (regarding Resources), also generally disagreed with Q.21-27 (regarding Policy Input).
- iv. A moderate to a strong positive relationship between PC res2 and PC inp2, indicating that those senior CSC officials that disagreed with Q.5, Q.6 and Q.11, and agreed with Q.7-10, also generally agreed with Q.21-25 (regarding Policy Input).

2. Experimental Design, Materials and Methods

2.1. Measurement

The scales of measurement in the survey were based around ordinal and interval scales where numerical values are assigned to a measurement [1]. Using Likert scales, applied previously in research on community sport [2], responses ranged from Strongly Agree, Agree, Undecided, Disagree to Strongly Disagree. After establishing background details, essential data collected was based upon twenty-three statements, as per Table 1.

To offer the potential to elicit themes in the responses in Table 1, the statements were grouped into four sections, namely: Resources (Q. 5-11); Policy Implementation (Q. 12-16); Policy Issues (Q. 17-20); and Policy Input (Q. 21-27). This grouping facilitated the statistical analysis, particularly identifying the correlations between each statement and theme. The questionnaire was sent to a total of 202 CSCs across the ten sports, with participants who completed the entire questionnaire numbering seventy-six ($n=53$) with a response rate of 26.2%.

3. Data Collection

In three specific areas – resources, policy, and advocacy – the objective for the data collection process was to recognize and isolate specific variables concerning community sport, then seek correlation and understand relationships between those variables to address the RQ. Occurring in 2019, the research was conducted by way of an online questionnaire and all correspondence and ethics approval was provided by the lead researcher's academic institution. An initial pilot survey was tested amongst seven academics with comments applied to improve the survey

¹ The direction of the relationship outlined may appear counterintuitive to the description that follows. This is due to the direction of the relationship between each PC and the statements themselves. For example, if a PC has a positive relationship with the statements as described in the section above, this indicates a higher PC value represents 'agree' and 'strongly agree' responses, whereas a lower PC value represents 'disagree' and 'strongly disagree' responses. On the other hand, if a PC is negatively correlated with a statement, this indicates that a higher PC value represents 'disagree' and 'strongly disagree'. Therefore, if there is a negative or positive relationship between two PCs, the direction of each statement with the respective PCs is considered to describe a *real-world* relationship.

format. Subsequently, the participants were sent an online questionnaire. Participants were provided with a uniform resource locator (URL) that incorporated Qualtrics and was labelled with the relevant university logo.

3.1. Participants

Survey participants were sourced by way of a targeted approach focussing on senior CSC officials (e.g., President/Secretary/Treasurer) who were purposively sampled from publicly available data on CSC websites of ten sports in the Illawarra region, as follows: *Australian Rules Football (AFL), Athletics, Cricket, Rugby League, Lifesaving, Netball, Soccer, Swimming, Touch Rugby, and*

Table 1
Questionnaire statements.

Resources (Q. 5 – 11)	<p>Q5 Adequate funding has been allocated to my club.</p> <p>Q6 My club has sufficient resources to achieve NSO/government policy objectives for community sport.</p> <p>Q7 Access to improved management systems would benefit my club.</p> <p>Q8 My club has resources it could share with other clubs (this might include clubs in the same sport but equally could apply to local community clubs in other sports).</p> <p>Q9 Other clubs have resources they could share with my club (this might include clubs in the same sport but equally could apply to local community clubs in other sports).</p> <p>Q10 Shared resources would enhance relationships between clubs (this might include clubs in the same sport but equally would apply to local community clubs in other sports).</p> <p>Q11 My club uses a performance management system to ensure that it adheres to government and NSO sport policy.</p>
Policy Implementation (Q. 12 – 16)	<p>Q12 The NSO provides clear guidelines on how national plans will be implemented locally.</p> <p>Q13 The NSO provides a clear management structure for policy implementation at local level.</p> <p>Q14 The NSO and my club cooperate effectively to improve efficiency in relation to policy implementation.</p> <p>Q15 NSOs, government and my club communicate on a regular basis to ensure coordinated delivery of activities designed to grow participation in sport.</p> <p>Q16 There is a clear commitment to a policy delivery system for community sport from national to local level.</p>
Policy Issues (Q. 17 – 20)	<p>Q17 There are frequent differences between my club and the NSO regarding the priorities for local level community sport.</p> <p>Q18 The NSO prioritizes performance sport over programs to increase participation levels in sport.</p> <p>Q19 Policy implementation is the sole responsibility of my club.</p> <p>Q20 My NSO provides clear channels for policy dispute resolution.</p>
Policy Input (Q. 21 – 27)	<p>Q21 There are mechanisms in place for feedback and review between the NSO and my club.</p> <p>Q22 It is important for clubs to play an active role in shaping <u>national</u> sport-related policy.</p> <p>Q23 My club plays an active role in shaping <u>national</u> sport-related policy.</p> <p>Q24 It is important for clubs to play an active role in shaping <u>local</u> sport-related policy.</p> <p>Q25 My club plays an active role in shaping <u>local</u> sport-related policy.</p> <p>Q26 My club would benefit from support from a local advocacy group that is independent from the NSO and government.</p> <p>Q27 The potential for funding is the primary factor that might influence my club supporting a local advocacy group.</p>

Rugby Union. Of these, five sports – AFL, Cricket, Netball, Soccer and Touch Rugby – are among the most popular in Australia based on *per capita* participation rates and rank in Australia's top twenty physical activities [3]. Although a non-exhaustive list in terms of representing community sport in the Illawarra region (e.g., golf and tennis were not included, nor were physical activities like walking or gym/fitness), the sports selected were based on (a) being team sports (or sports where teams were often involved) and (b) ordinarily requiring a high number of officials/organizational effort to host a competition or event.

Representing just over 26% of senior CSC officials to whom the questionnaire was sent, based on average trends in organizational and academic research [4], the response rate was sufficient for the objectives of this paper. The sample size was impacted positively with one follow-up email to senior CSC officials requesting completion of the survey, a measure that helped ensure a good response rate, considered important for statistically significant research [5]. The result provided a position of statistical significance by virtue of a sample size that was large enough to be confident of detecting a beneficial effect or relationship between the variables.

3.2. Statistical analysis

The approach taken to the analysis is descriptive, and the responses from the senior CSC officials were evaluated once. In addressing the RQ, the objective was to outline relationships between variables to determine that a valid estimate of a generalized or pervasive connection between variables was achieved. To identify relevant correlations between variables, confidence intervals and Gaussian copula graphical models (GCGMs) – which accommodate and can be applied to a broad range of data types [6,7] – were used. While frequently associated with finance, science and medicine, the application of copulas and GCGMs is increasingly common in research in sport where there are multiple variables and relationships involved [8,9]. GCGMs have been used in, for example, analysis in soccer and across physical education in general [10,11].

Statistical analysis was conducted in RStudio, applying R version 3.6.1 [12]. The percentage, along with 95% confidence intervals for the percentage, of participants who responded to each statement with Strongly Agree, Agree, Undecided, Disagree and Strongly Disagree, was calculated using the *'MultiNomCI'* function from the *'DescTools'* package [13]. To determine the raw and partial correlations between each statement, the *'cgr'* function from the *'ecoCopula'* package, combined with multiple univariate Cumulative Link Models (CLM) using the *'clm'* function from the *'ordinal'* package [14] were pooled using the *'manyany'* function from the *'mvabund'* package [15]. These models were matched with Sport (Q.1) and Position (Q.2) as covariates. This method has been used previously to distinguish direct associations from indirect mediator effects in multivariate data [7].

The research design was illustrative and sought to establish associations between variables based on the responses from senior officials of CSCs. GCGMs were used to uncover and illustrate conditional relationships among CSCs. Copula models are so named because they couple a multivariate distribution with any set of marginal distributions [7]. GCGMs have significant potential for the analysis of multivariate data [16] and the ability to reveal conditional dependence relationships between subjects [17]. Principal components analysis (PCA) was conducted using the *'princals'* function from the *'Gifi'* package for exploratory and data reduction purposes [18] to represent each set of themed statements by a smaller number of 'artificial' variables that would explain as much variation in the original statement as possible. Following the PCA, multiple General Additive Models (GAMs) were set using a combination of the *'gam'* function from the *'mgcv'* package [19] and the aforementioned copular graphical modelling methods. This was performed to determine the conditional relationships between the new artificial variables.

Ethics Statement

The research was conducted by way of an online questionnaire and informed consent was obtained from all survey participants. Ethics approval was granted by the author's academic institution, the University of Canberra, under protocol number 20180310.

CRedit Author Statement

Conceptualization, Investigation, Methodology, Data curation, Formal analysis, Software, Validation, Visualization, Original Draft Preparation, Writing – review & editing.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

CCGM Data (Original data) (Mendeley Data).

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.dib.2022.108111](https://doi.org/10.1016/j.dib.2022.108111).

References

- [1] J. Skinner, A. Edwards, B. Corbett, *Research methods for sport management*, Routledge (2014).
- [2] P. Wicker, K. Filo, G. Cuskelly, Organizational resilience of community sport clubs impacted by natural disasters, *J. Sport Manag.* 27 (6) (2013) 510–525.
- [3] SportAus, Australia's top 20 sports and physical activities revealed. https://www.sportaus.gov.au/media-centre/news/australias_top_20_sports_and_physical_activities_revealed, (2019).
- [4] Y. Baruch, B.C. Holtom, Survey response rate levels and trends in organizational research, *Human relations* 61 (8) (2008) 1139–1160.
- [5] E. Babbie, *The practice of social research*, Nelson Education Ltd, Wadsworth, 2010.
- [6] G.C. Popovic, F.K. Hui, D.I. Warton, A general algorithm for covariance modeling of discrete data, *J. Multivariate Anal.* 165 (2018) 86–100.
- [7] G.C. Popovic, D.I. Warton, F.J. Thomson, F.K. Hui, A.T. Moles, Untangling direct species associations from indirect mediator species effects with graphical models, *Methods Ecol. Evol.* 10 (9) (2019) 1571–1583.
- [8] M. Atkinson, *Key concepts in sport and exercise research methods*, Sage (2011).
- [9] J. Lennartsson, Probabilistic modeling in sports, finance and weather, (2014).
- [10] B. Eberth, M.D. Smith, Modelling the participation decision and duration of sporting activity in Scotland, *Econ. Modell.* 27 (4) (2010) 822–834.
- [11] M. Tavassolipour, M. Karimian, S. Kasaei, Event detection and summarization in soccer videos using bayesian network and copula, *IEEE Trans. Circuits Syst. Video Technol.* 24 (2) (2013) 291–304.
- [12] R.C. Team, A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; (2012), URL <https://www.R-project.org> (2019).

- [13] A. Signorell, K. Aho, A. Alfons, N. Anderegg, T. Aragon, A. Arppe, A. Baddeley, K. Barton, B. Bolker, H. Borchers, DescTools: Tools for descriptive statistics. R package version 0.99.28, R Found. Stat. Comput., Vienna, Austria (2019).
- [14] R. Christensen, ordinal-regression models for ordinal data. R package version 2019.4-25, Retrieved May 1 (2019).
- [15] Y. Wang, U. Naumann, D. Eddelbuettel, J. Wilshire, D. Warton, J. Byrnes, R. dos Santos Silva, J. Niku, I. Renner, S. Wright, mvabund: Statistical methods for analysing multivariate abundance data. R package Version 4.0.1, (2019).
- [16] M.J. Anderson, P. de Valpine, A. Punnett, A.E. Miller, A pathway for multivariate analysis of ecological communities using copulas, *Ecol. Evol.* 9 (6) (2019) 3276–3294.
- [17] N. Meinshausen, P. Bühlmann, High-dimensional graphs and variable selection with the lasso, *Ann. Stat.* 34 (3) (2006) 1436–1462.
- [18] P. Mair, J. De Leeuw, P. Groenen, Gifi: Multivariate analysis with optimal scaling, R package version 0.3-7/r266 (2017).
- [19] S.N. Wood, *Generalized additive models: an introduction with R*, CRC press, 2017.