Slov Vet Res 2023; 60 (Suppl 25): 307–15
DOI 10.26873/SVR-1608-2022

Original Research Article
Veterinary Medicine and The One Health Concept

A COMBINED APPROACH OF MULTIPLE CORRESPONDENCE ANALYSIS AND HIERARCHICAL CLUSTER ANALYSIS FOR PROFILING RELATIONSHIPS AMONG CATEGORICAL RISK PREDICTORS: A BLUETONGUE CASE STUDY

Iman E. El-Araby[1], Sherif A. Moawed[2], Fardos A.M. Hassan[1], Hagar F. Gouda[1]*

[1]Department of Animal Wealth Development, Faculty of veterinary medicine, Zagazig University, 44511, [2]Department of Animal Wealth Development, Faculty of veterinary medicine, Suez Canal University, 41522, Egypt

*Corresponding author, E-mail: hagarfathy@zu.edu.eg

Abstract: Bluetongue (BT) is a non-contagious virus in the Reoviridae family that infects both wild and domestic animals. It causes economic losses and reduces infected animals' production and reproduction. This study aims to show the utility of MCA and HCCA in identifying relationships between categorical variables used as risk factors for Bluetongue disease. Six categorical variables (age, sex, season, species, locality, and BT serotyping) of 233 apparently healthy animals were screened for studying BT. Based on visualized information of MCA and HCA among variable categories, disease profiles were identified. The first two MCA dimensions retained up to 27% of the total inertia contained in the data. The positive BT results, summer, and old animals categories were loaded in the first dimension, while negative cases, Al-mounfia and winter categories were related to the second dimension. HCA identified three clusters. Cluster 1 was characterized by frequent and largely exclusive seronegative BT animals 91.67 % of animals in the cluster were seronegative, negative BTV category is the most important and related to cluster 1 with positive v-test=8.75. Cluster 3 can named a cluster of seropositive BT, up to 88% of cases were seropositive. We can conclude that seropositive BT is associated with summer and old age categories, whereas seronegative BT is associated with young age and winter categories, and thus MCA and HCA provide convenient and easy-to-interpret analytical tools for assessing categorical data relationships.

Key words: categorical data; multiple correspondence; inertia; factominR; hierarchical cluster

Introduction

Bluetongue virus (BTV) is a non-contagious vector-borne infection of both domestic and wild animals that belongs to the genus Orbivirus and the family Reoviridae. Bluetongue virus infection is spread to susceptible animals such as sheep, goats, and deer by mosquitoes, ticks, and species of biting midges (Culicoides) (1). Reduction in weight, low milk yield, wool breakage, and other economic losses are caused by this disease (2), and also can cause a high mortality rate (70%) in highly susceptible sheep (3).

Data in epidemiological studies almost collected in vast volumes using surveys and the response pattern involves many categories which may be either binary, ordinal, or nominal. Researchers are frequently interested in investigating the links between such sets of category variables. Consider doing independent chi-square tests for each pair of variables, or, in the case of binary or ordinal data, viewing a correlation matrix of the bivariate associations. However, for a high number of categorical variables, this paired technique would become tedious and the findings would be

Received: 17 September 2022
Accepted for publication: 14 November 2022

difficult to describe. More considerably, such a technique would only indicate the existence of a relationship but not which response categories are associated (4).

Multiple correspondence analysis (MCA) is one of a family of descriptive approaches and an extension of correspondence analysis that allows studying the pattern of relationships between numerous categorical dependent variables. CA's multivariate extension is used to analyze tables with three or more variables. Furthermore, MCA can be thought of as an expansion of principal component analysis for categorical variables, revealing patterns in large data sets. MCA aids in the distinct description of patterns of relationships using geometrical methods by situating each variable/unit of analysis as a point in a low-dimensional space. MCA can be used to map variables as well as individuals, allowing the creation of complex visual maps whose structure can be analyzed. Furthermore, this technique has the ability to connect both variable-centered and case-centered approaches (5).

The MCA technique, as opposed to the orthogonalization technique that underpins PCA, employs a distance measure. MCA converts the relationship between discrete variable categories into coordinates in a multidimensional space. It assigns scale values to the discrete variable categories and maximizes the variance of those scores to determine the correlations between the variables and the proximity of subjects. Points pointing in the same direction as the origin are strongly linked. The mean is represented by points near the origin, whereas points farther from the origin depart from the mean (6).

Clustering is a type of unsupervised learning job that discovers underlying structures in unlabeled data. Those are divided into homogeneous groups or clusters, with intracluster items having high resemblance while being quite distinct to objects in other clusters. Over the years, many clustering approaches have been suggested and implemented. Clustering methods are classified into two types: hierarchical clustering and partitional clustering. While hierarchical clustering uses agglomerative or divisive algorithms to build a hierarchy of partitions (i.e., a dendrogram) across the dataset, partitional clustering usually assumes a fixed number of clusters and strives to maximise homogeneity within the clusters (7).

Hierarchical cluster algorithms portray data as a tree of nodes, with each node representing a possible data categorization. Hierarchical algorithms can be used to cluster categorical data in two different manners: an agglomerative (bottom-up) and divisive (top-down). The latter, on the other hand, is less common. The agglomerative algorithm's core idea is to use a similarity metric to gradually assign objects to tree nodes. The fundamental drawback of hierarchical clustering is its poor pace. Another issue is that the clusters may merge, causing these methods to cause information distortion (8).

The chi-square test or logistic regression is frequently used in epidemiologic studies to assess the relationship of qualitative risk factors. MCA is not as widely used as these methods, despite its importance and convenience in representing relationship patterns in data. The purpose of this study is to use MCA and HCCA to describe and assess the relationship between bluetongue risk factors in five Egyptian governments and bluetongue seropositivity in small ruminants. Using the MCA graphical plot and HCCA clusters, it is simple to identify, if possible, any hidden patterns of animals based on these risk factors.

Material and methods

Source of data

The data were obtained from a previous epidemiologic study on 233 apparently healthy animals: 125 sheep (42 males and 83 females) and 108 goats (47 males and 61 females). The animals were screened for bluetongue seropositivity using competitive ELISA from April 2018 to March 2019. The description of the data is shown in table 1.

Methodology and model

There are K variables, each with a level of Jk, and the total number of Jk is equal to J. There are also as many observations as I. X represents the $I \times J$ matrix. When a correspondence analysis is performed on the indicator matrix, two groups of factor scores are produced: one for rows and one for columns. In general, these factor scores are adjusted so that the variance equals the appropriate eigenvalues.

Table 1: Epidemiologic qualitative variables for studying bluetongue

variables	Categories
1. Species	Sheep or goat
2. Age	Young (6- < 18 months), moderate (18- < 36 months), and old (≥ 36 months)
3. Sex	male and female
4. locality	Al-Sharkia, Al-Monfia, Al-Menia, Al-Giza, and Al-Suis
5. Season	winter, spring, summer, and autumn
6. Bluetongue	Positive or negative serotyping.

The probability matrix $Z = N\text{-}1X$ is calculated as the initial stage in the investigation. Where N is the total value of the arrangement's matrix, r is the vector of the total row Z (i.e., r = Z1, where 1 is the vector of 1), and c is the vector of the whole column. Dc = diag {c} and Dr = diag {r}. The factor score is calculated by decomposing a single number in the following equation:

$$D_r^{-\frac{1}{2}}(Z - rc^T) \qquad D_c^{-\frac{1}{2}} = P\Delta Q^T$$

(Δ is a diagonal matrix of single values, and Λ=Δ2 is a matrix of eigenvalues). The row and column factor scores (respectively) are obtained by means of the following equation:

$$F = D_r^{-\frac{1}{2}} P\Delta \qquad G = D_c^{-\frac{1}{2}} Q\Delta$$

Distance squared (x2) of rows and columns can be denoted in the following equation:

$$d_r = \text{diag}\{FF^T\} \quad d_c = \text{diag}\{GG^T\} \quad (9).$$

Hypothesis and assumptions:

MCA is assumption-free, and when working with categorical data, it can represent both linear and non-linear connections equally well. These are significant advantages over more traditional methodologies, which may require linear correlations between variables and a prior hypothesis is about likely variable interactions (10).

Validity of MCA:

The validity of MCA was assessed according to Rodriguez-Sabate *et al.* (10) by:

1. Inertia: Which depicts data dispersion around their center of gravity G (or centroid) and is used as a measure of information. The term inertia is employed by analogy with the notion of "moment of inertia" in practical mathematics, which stands for the integral of mass times the squared distance between the centroids. Where;

$$\Delta^2 = \sum_i p_i + d_i^2$$

p_i is the marginal relativefrequency (or mass) of row i, and is the chi-square distance between the row i's profile and the average row profile, All aforementioned equations are analogous for column profiles (11). $d_i = \sqrt{\sum_j \left(\frac{(p_{ij}-p_{+j})^2}{p_{+j}}\right)}$

2. Total inertia which represents the inertia of all the dimensions analyzed, and ranged from 0 which denotes no information to 1 which denotes all available information.
3. Contribution of each variable in extracted dimensions, and ranged from 0 to 1.
4. The quality of a variable (cosine2) Cosine2 is a value represents the squared cosine value of the angle created by the point with the specified dimension. The closer the number is to one, the better the representation of the variable in the computed dimensions.
5. Eigenvalues Indicate the relative weight of each dimension in relation to total inertia (it is normalized to 1 which represents all the information of all the variables in all the dimensions). The first dimension always had the largest eigenvalue, which decreased steadily through the remaining dimensions. This variable (together with cumulative inertia) is typically used to determine the number of dimensions to include in the MCA. As a result, dimensions with eigenvalues less than 0.05 are commonly ignored.

Cluster analysis

Clustering is defined by two functions: the distance function and the linkage function. The distance function calculates the distances between the points, while the linkage function calculates the distance between clusters. Clustering outcomes frequently differ depending on the functions used. The distance between two points is defined as:

$$\varphi_{j,k} = \|p_j - p_k\| = \sqrt{(x_j - x_k)^2 + (y_i - y_k)^2}$$

Which is the Euclidean distance. The clusters depend on the choice of a clustering distance $d \geq 0$. Then, if two points satisfy $\varphi_{j,k} \leq d$, they are in the same cluster. Next, let A and B be two clusters containing points a_α and b_β, then the distance between two clusters is defined by:

$$\varphi(A, B) = \min_{\alpha,\beta} \varphi(a_\alpha, b_\beta)$$

which is known as single-linkage merge criterion (12, 13).

Ward's linkage is most commonly used with the Euclidean squared distance measure (14). At each step, Ward's linkage combines the two clusters, C_m and C_k, into one cluster C_l, that minimizes the total within-cluster squared error. The within-cluster squared error $S(C_l)^2$ of the lth cluster is defined as:

$$S(C_l)^2 = \sum_{i \in C_l} \sum_{p=1}^{P} (x_{ip} - \frac{1}{|C_l|} \sum_{j \in C_l} x_{jp})^2$$

where $|C_l|$ is the number of observations in cluster C_l, P is the number of variables measured for each observation, x_{ip} and x_{jp} denote the pth variable corresponding to observations i and j.

The total within-cluster squared error is the sum of within-cluster squared error over all K clusters (14, 15):

$$\sum_{l=1}^{K} S(C_l)^2$$

Software used

All analyses were conducted by SPSS version 25 (Armonk, NY: IBM Corp) and RStudio (16) using R (17).

Results

The results show that ten dimensions were extracted and described 100% of information of data. The first two dimensions 1 and 2 are sufficient to retain 27% of the total inertia (variation) contained in the data as shown in fig. 1. In table 2. The dimension's eigenvalue that is a measure

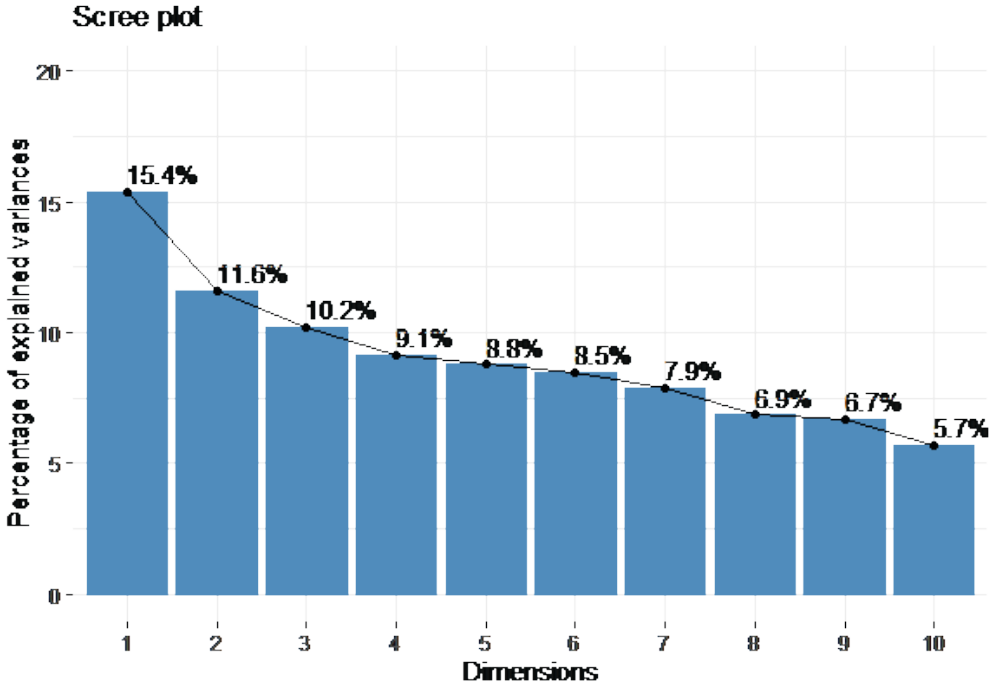

Figure 1: Scree plot for the main 10 extracted dimensions

Table 2: Discrimination and contributions of variables for the 1st and 2d dimensions

	Dimension 1			Dimension 2		
	Discrimination (R^2)	**contribution**	**Cos2**	**Discrimination (R^2)**	**contribution**	**Cos2**
Species	0.152	8.24%	0.3	0.131	9.43%	0.26
Sex	0.163	8.83%	0.32	0.092	6.63%	0.18
Age	**0.344**	18.65%	0.52	0.117	8.41%	0.17
Locality	**0.350**	18.94%	**0.53**	**0.408**	29.26%	0.45
season	**0.341**	18.45%	**0.47**	**0.474**	34.04%	0.63
BTV result	**0.496**	**26.87%**	1	0.170	12.22%	0.34
eigenvalues	0.31			0.23		
percentage of variance	15.39%			11.61%		

Cos2 denotes the squared cosine; Discrimination measures quantify the variance in each indicator.

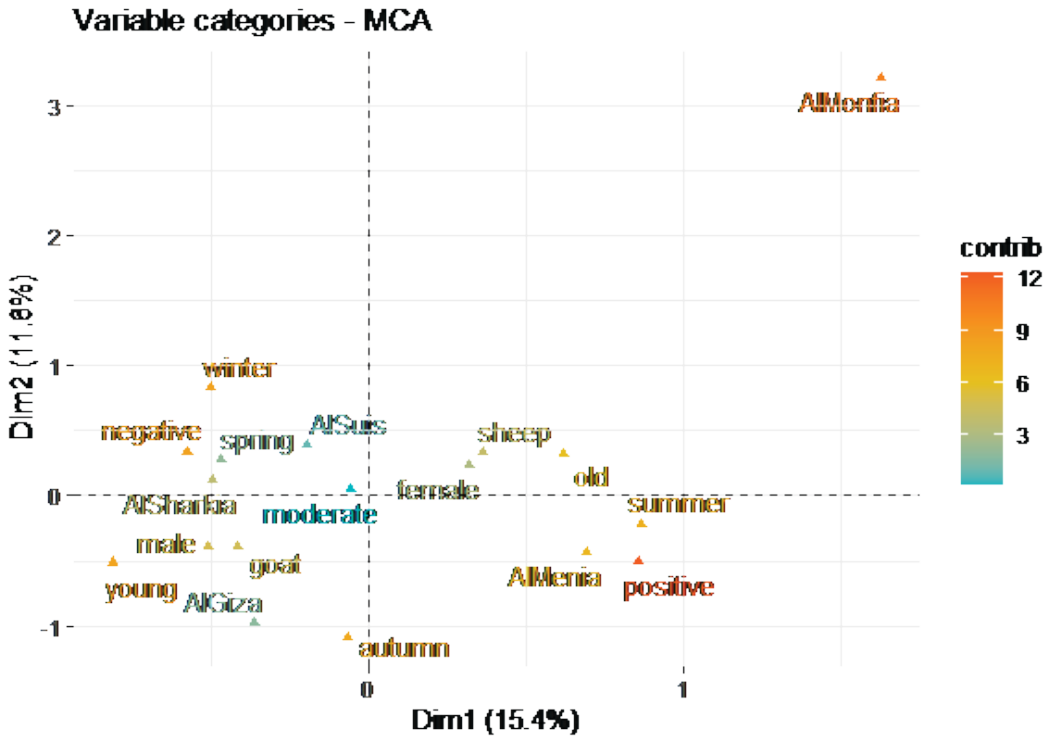

Figure 2: variable contributions on dimensions. Variable categories with low contribution values will be colored in "white", variable categories with mid contribution values will be colored in "blue", variable categories with high contribution values will be colored in "red"

of the amount of variance for which it accounts. Dimension 1 accounts for 0.31 of variance with percentage of 15.39% and dimension 2 for 0.23 of variance with 11.61%.

Table (2) and figure (2) help to find the variables that are most closely related to each dimension. The squared correlations between variables and the dimensions are used as coordinates. BTV result, locality, age, and season represent the indicators that most discriminating and making the largest contribution to dimension one. While locality and season are the most discerning indicators and contribute the most to dimension two. The squared cosine (cos2), which gauges the degree of association between different variable categories and a particular axis, measures discrimination (dis.), contribution (ctr.), and representation quality table (2). The most contributing and discriminating variables for the first dimension are BTV result (dis. = 0.496, ctr. = 26.87%), locality (dis. = 0.35, ctr. =18.94%), age (dis. = 0.344, ctr. =18.65%) and season (dis. = 0.341,

ctr. =18.45%). The cos2 of variable categories that can be extracted as in table (2). If a variable category is well represented by two dimensions, the sum of the cos2 is closed to one. The result BTV variable is the variable with highest cos2 = 1, then locality=0.53, age= 0.52, and sex=0.47. For the second dimension the most contributing and discriminating variables are season (dis. = 0.474, ctr. = 34.04%) and locality (dis. = 0.408, ctr. = 29.26%).

Table (3) lists all of the BTV-contributing factors in this study in descending order of significance, considering the coefficient of determination (R2) and the P value of the overall test (F-test). The R^2 value ranges from 0 to 1, with 0 being no relationship and 1 being a very strong relationship between the qualitative variable and the MCA dimension. The presence of sex categories (male and females) are somewhat far from positive and negative categories of bluetongue results, which is also confirmed by the significance of variables.

Whereas no other correlation was found as shown in table (4). Only correlations above 0.30 were considered to have meaningful practical significance. Table (5) shows the contribution of variables' categories for each of the first two dimensions. For dimension 1; the highest contributions and cos2 with positive coordinate (coord.) were for positive BTV (ctr. =16.03, cos2. =0.5, coord. = 0.86), summer season (ctr. = 11.98, cos2. =0.31, coord.= 0.86), old (ctr. = 8.73, cos2. =0.26, coord.= 0.62), and Al-Minea (ctr. = 8.55, cos2. =0.24, coord.= 0.69). When looking at the second dimension; the highest contributions and cos2 with positive coordinate were for Al- Monufia (ctr. = 19.06, cos2. =0.27, coord.= 3.21), winter

Table 3: Significance of test results for key BTV contributing factors in the top 2 dimensions

	R^2	P-value		R^2	P-value
Dim.1			**Dim.2**		
Result	0.496	< 0.00001	**season**	0.47	< 0.00001
Age	0.344	< 0.00001	**Locality**	0.41	< 0.00001
Season	0.341	< 0.00001	**Result**	0.17	< 0.00001
locality	0.3499	< 0.00001	**Species**	0.13	< 0.00001
sex	0.16	< 0.00001	**Age**	0.12	< 0.00001
species	0.15	< 0.00001	**sex**	0.09	< 0.00001

In dimension 1, age correlated (transformed variables) significantly with species (r = 0.26, r < 0.001), locality (r = 0.207, r < 0.001), and season (r = 0.142, r < 0.05); sex correlated with locality (r = 0.222, r < 0.001), locality correlated with season (r = 0.201, r < 0.01) dissimilar correlations were found for dimension 2, just age correlated with species (r = 0.154, r < 0.05) and sex with season (r = 0.13, r < 0.05).

Table 4: variables relationship on dimensions1&2

	Species	Sex	Age	Locality	Season
Dim.1					
Species	1				
Sex	0.102^{NS}	1			
Age	0.26**	0.101^{NS}	1		
Locality	0.099^{NS}	0.222**	0.207**	1	
season	0.045^{NS}	0.125^{NS}	0.142*	0.201** (0.002)	1
Result					
Dim.2					
Species	1				
Sex	-0.102^{NS}	1			
Age	0.154*	0.036^{NS}	1		
Locality	0.096^{NS}	0.057^{NS}	0.051^{NS}	1	
season	-0.013^{NS}	0.13*	0.012^{NS}	0.126^{NS}	1
Result					

NS: non-significant correlation P > 0.05; * significant correlation P<0.05; ** highly significant correlation P < 0.01

Table 5: Contributions of variables' categories for each dimension of top 2 dimensions

	contributions	Cos2	coordinates	contributions	Cos2	coordinates
	Dim.1			Dim.2		
Species						
Goat	4.42	0.15	**-0.42**	**5.06**	**0.13**	**-0.39**
Sheep	3.82	0.15	**0.36**	**4.37**	**0.13**	**0.34**
Sex						
Female	3.37	**0.16**	**0.32**	**2.53**	0.09	**0.24**
male	5.46	**0.16**	**-0.51**	**4.10**	0.09	**-0.39**
Age						
Young	10.22	**0.26**	**-0.82**	**5.37**	**0.1**	-0.51
Moderate	0.06	**0.00**	**-0.06**	**0.06**	**0.00**	0.05
old	8.37	**0.26**	**0.62**	**2.98**	**0.07**	0.32
Locality						
Al-Giza	0.37	**0.01**	**-0.37**	**3.57**	**0.05**	-0.98
Al-Minea	8.55	**0.24**	**0.69**	**4.54**	**0.09**	-0.44
Al-Monfia	3.70	**0.07**	**1.63**	**19.06**	**0.27**	3.21
Al-Sharqia	6.01	**0.20**	**-0.5**	**0.49**	**0.01**	0.12
Al-Suiez	0.31	**0.01**	**-0.2**	**1.60**	**0.03**	0.39
Season						
Winter	3.89	**0.1**	**-0.5**	**13.96**	**0.27**	0.83
Spring	2.53	**0.06**	**-0.47**	**1.16**	**0.02**	0.28
Summer	11.98	**0.31**	**0.86**	**0.99**	**0.02**	-0.22
autumn	0.05	**0.00**	**-0.07**	**17.93**	**0.32**	-1.09
BTV result						
Positive	16.03	**0.5**	0.86	**7.92**	**0.17**	**-0.5**
negative	10.84	**0.5**	-0.58	**4.93**	**0.17**	**0.34**

Cos2: is the quality or squared correlations of each dimension

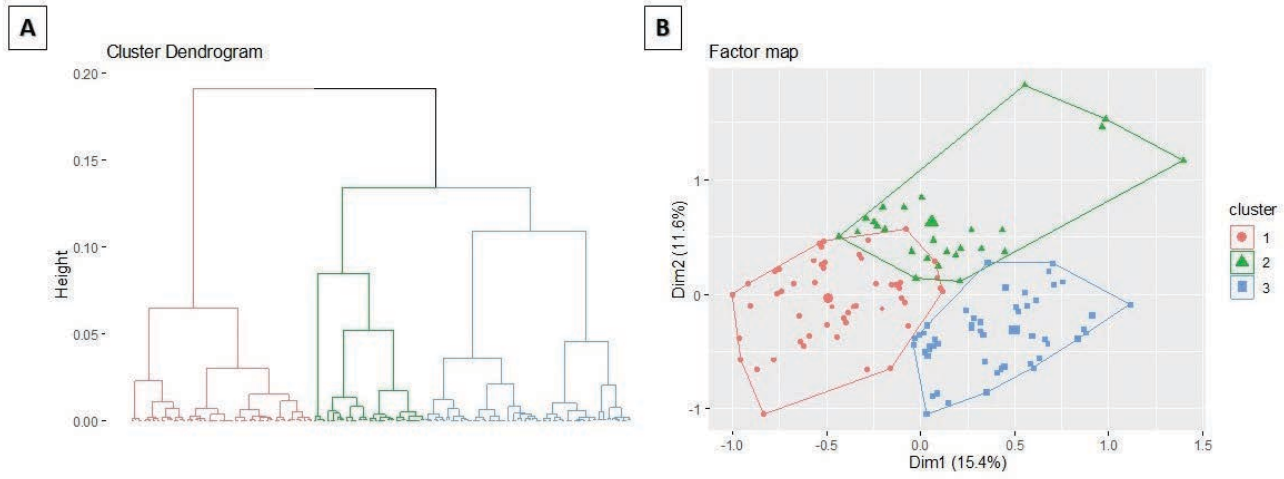

Figure 3: (A): Dendrogram of correspondence scores revealing the three clusters. The hierarchical tree suggests a clus-tering into three clusters. Figure (B): the three clusters on factor map

(ctr. = 13.96, cos2. =0.27, coord.= 0.83), negative BTV results (ctr. = 4.93, cos2. =0.17, coord.= 0.34), and sheep (ctr. = 4.37, cos2. =0.13, coord.= 0.34).

Data from all MCA dimensions were utilized in the HCCA to determine the animals' profiles and clusters. The results show three clusters figure 3. "BTV result" and "season" are the variables that best describe the division into three clusters. Only the variables' categories whose p-value is less than 0.02 are used. Only the categories whose p-value is less than 0.02 are used. Cluster 1 (negative BTV young aged animals): This group was characterized by frequent and largely exclusive seronegative BTV animals 91.67% of animals in the cluster are negative, negative BTV category is the most important and related to cluster 1 with positive v-test=8.75. 71.88% of animals in this cluster

are from Al-Sharkia governorate and 52% are young age. Cluster 2: 75% of cases recorded in winter (v-test= 7.64), up to 40% in Al-Suis, and 62% are old age. Cluster 3 (seropositive BTV): Out of all animals in this cluster 87.64% are positive BTV, 62.92% of cases are recorded from Al-Minea governorate, 55% and up to 35% are in summer and autumn respectively. 73% are females, and 52% are old.

Discussion

The study investigated the relationship among six qualitative variables involving 18 categories using MCA and HCA. MCA is appealing because it provides a bi plot depiction of both variables and variables' categories, which is not available in many qualitative data relationship analytical tools. HCA based on data of MCA revealed that data is well separated by season and BTV result into three clusters. The seropositive BT category was related to summer, and Al-Minea categories this finding is in contrast with Malek and Abou EL-wafa (18) who recorded more seropositive of BTV in non-hot months compared to hot months. Old age, female, and sheep categories are associated and all are grouped and associated with positive BTV category. These findings are consistent with (19) who reported that seroprevalence increases with age, probably a reflection of increased duration of exposure. In a previous study (20) they found that the seroprevalence rates were increased with increase of age in sheep and goats. However, in a seroprevalence study of Mozaffari and Khalili (21) in the south-east of Iran the results showed that seroprevalence rates were decreased with the increasing of age in sheep. The left side of plot mainly obtain negative BTV results, which associated with winter and spring. The governorates that showed more negative BTV were Al-Suiez, Al-Sharqia, and Al-Giza, which is in contrast with (22) who reported the two governorates with the highest prevalence with Beni-seuif governorates. The categories that show a degree of association to negative BTV are male, goat, and young age categories. The bottom-left quadrant shows the association between male, young, goat, and Al-Giza which appears more related to negative results compared to positive results. The presence of sex categories (male and females) are somewhat far from positive and negative categories of bluetongue results, which is also confirmed by the significance of variables in table (7). Sex demonstrated lower significance and correlation to the dimensions. The result is in line with (20) who applied chi-square and found non-significant effect of sex to seropositivity of BTV. However, Malek and Abou EL-wafa (18) reported that sex is significantly related to seropositivity and BTV is more recorded in females.

Conclusion

In large-scale qualitative data sets with more than two variables or survey-collected epidemiologic data, multiple correspondence analysis is a useful technique for identifying relationships. Additionally, it is interpretable for assessing relationship patterns and free of assumptions. By grouping data points together based on their homogeneity, HCA can be used for categorical data after MCA to display data patterns and reveal the underlying structure of BT data (positive or negative serotyping) in relation to other variables and/or categories.

Acknowledgements

The authors would like to acknowledge and thank Hend El-Mohamedy for her help in availability of data.

References

1. Bouwknegt C, van Rijn PA, Schipper JJ, et al. Potential role of ticks as vectors of bluetongue virus. Exper Appl Acarol 2010; 52(2): 183–92.

2. Wilson AJ, Mellor PS. Bluetongue in Europe: past, present and future. Philosph Trans Royal Soc Biol Sci 2009; 364(1530): 2669–81.

3. Breard E, Hamblin C, Hammoumi S, et al. The epidemiology and diagnosis of bluetongue with particular reference to Corsica. Res Vet Sci 2004; 77(1): 1–8.

4. Nagpaul P. Guide to advanced data analysis using IDAMS software. New Delhi: United Nations Educational, Scientific and Cultural Organization. 1999.

5. Ayele D, Zewotir T, Mwambi H. Multiple correspondence analysis as a tool for analysis of large health surveys in African settings. Afr Health Sci 2014; 14(4): 1036–45.

6. Dungey M, Doko Tchatoka F, Yanotti MB. Using multiple correspondence analysis for finance: A tool for assessing financial inclusion. Int Rev Finan Anal 2018; 59: 212–22.

7. Nguyen HH. Clustering categorical data using community detection techniques. Comput Intell Neurosci. 2017; 2017.

8. Gevorgyan RA, Hakobyan YB. A matching based clustering algorithm for categorical data. arXiv preprint arXiv:181203469. 2018.

9. Abdi H, Valentin D. Multiple Correspondence Analysis. Encyclopedia of Measurement and Statistics. 2007.

10. Rodriguez-Sabate C, Morales I, Sanchez A, Rodriguez M. The Multiple Correspondence Analysis Method and Brain Functional Connectivity: Its Application to the Study of the Non-linear Relationships of Motor Cortex and Basal Ganglia. Front Neur 2017;11:345.

11. Clausen SE. Applied correspondence analysis: An introduction: Sage; 1998.

12. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM computing surveys (CSUR). 1999;31(3):264–323.

13. Jain AK, Dubes RC. Algorithms for clustering data: Prentice-Hall, Inc.; 1988.

14. StataCorp L. Stata multivariate statistics: reference manual. 17 ed: Stata Press Publication; 2021.

15. Ward Jr JH. Hierarchical grouping to optimize an objective function. J Amer Stat Ass 1963;58(301):236–44.

16. RStudio Team: integrated development for R. RStudio. Inc, Boston, MA. 2019.

17. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2021. Available from: https://www.R-project.org/.

18. Malek SS, Abou El-wafa S. High seroprevalence of bluetongue in sheep, goats and cattle in assiut governorate, Egypt. Kafrelsheikh Vet Med J 2016;14(1):285–96.

19. Radostits OM, Gay C, Hinchcliff KW, Constable PD. Veterinary Medicine E-Book: A textbook of the diseases of cattle, horses, sheep, pigs and goats: Elsevier Health Sciences; 2006.

20. Najarnezhad V, Rajae M. Seroepidemiology of bluetongue disease in small ruminants of north-east of Iran. Asian Pac J Trop Biomed 2013;3(6):492–5.

21. Mozaffari AA, Khalili M. The first survey for antibody against bluetongue virus in sheep flocks in southeast of Iran. Asian Pac J Trop Biomed 2012; 2(3): S1808–10.

22. Mahmoud M, Khafagi MH. Seroprevalence of bluetongue in sheep and goats in Egypt. Vet World 2014;7(4).