

CORRESPONDENCE ANALYSIS TO EXPLORE THE RELATIONSHIP BETWEEN CATEGORIES OF QUALITATIVE VARIABLES (STRUCTURAL CHROMOSOMAL ABERRATIONS-INFERTILITY IN EGYPTIAN BUFFALO)

FATMA D.M. ABDALLAH¹ AND ASMAA W. ZAGLOOL²

¹ Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Egypt.

² Genetics and Genetic Engineering Department, Faculty of Veterinary Medicine, Zagazig University, Zagazig, Egypt.

Received: 18 February 2025; **Accepted:** 25 August 2025

ABSTRACT

This study aims to explore the relationships between categorical variables using Correspondence Analysis (CA), a statistical technique designed to visualize and interpret associations in contingency tables. By applying CA, we seek to uncover underlying patterns in the structure of count data. The variables under study were structural chromosomal aberrations and some diseases related to infertility in Egyptian buffalo. Structural chromosomal aberrations were divided into 6 categories (gap, break, deletion, fragment, ring chromosome and centromeric attenuation). Groups of animal diseases were 9 categories (control, repeat breeder, anestrus, retained placenta, free-martin, vaginal prolapse, uterine prolapse, uterine torsion and habitual abortion). The Chi-square test of independence revealed a statistically significant association, indicating a relationship between chromosomal aberrations and infertility groups. CA further supported this association, with a total inertia of 0.178, suggesting that approximately 17.8% of the variation in the data is explained by the relationship between these two variables. Dimensions 1 and 2 captured most of the data structure, explaining 53.5% and 27.2% of the variance, respectively. Uterine torsion and abortion were highly contributed to explaining data variance in dimension 1. For dimension-2, uterine prolapse, uterine torsion, and control were also highly contributed to explaining variance. For the second variable, centromeric attenuation was highly significant for dimension-1 and fragment and centromeric attenuation were highly contributed to variance explaining than other structural chromosomal aberrations. Animal breeders can use CA techniques in their farms to facilitate understanding the pattern of their data and graphical representation of large data set.

Keywords: Correspondence analysis, chromosomal aberrations, graphical analysis, inertia, singular value decomposition.

INTRODUCTION

There are many statistical methods for analyzing and modelling contingency

tables data such as Chi-squared test, Cochran–Mantel–Haenszel statistics, correspondence analysis, G-test, linear discriminant analysis, log linear model, multinomial distribution, multiple correspondence analysis and many other tests.

Classical statistical methods for categorical data analysis, such as chi-square test, are

Corresponding author: Fatma D.M. Abdallah
E-mail address: moibho_stat2021@yahoo.com
Present address: Assistant Professor, Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Egypt.

not adequate in case of many qualitative variables (more than two) or multi-way contingency tables (Agresti, 1996).

Visualization of data or graphical representation of contingency tables is very important to understand data pattern and structure, especially in case of large datasets (Kumar & Nasser, 2016).

Correspondence analysis (CA) idea was first developed with Karl Pearson (1906). Then de Leeuw (1983) developed the coefficient of correlation in a two-way contingency table using linear regression.

Correspondence analysis is a descriptive multivariate exploratory technique used to analyze categorical data (Benzecri, 1992). Correspondence Analysis (CA) can be classified into two main types: Simple CA and Multiple CA. Simple CA is used to analyze the relationship between two qualitative variables, regardless of the number of categories, whereas Multiple Correspondence Analysis (MCA) is applied when there are three or more qualitative variables, and the interpretation becomes more complex (Abdi & Valentin, 2007). Correspondence analysis has become a widely accepted statistical method for exploring associations within complex categorical data. It is especially valuable in the social and biological sciences, where traditional modeling becomes challenging due to the large number of variables (Beh, 2007). Multiple Correspondence Analysis (MCA), in particular, is well-suited for large-scale qualitative datasets and survey-based epidemiological studies, providing interpretable insights into underlying patterns of association (El-Araby *et al.*, 2023).

Because of application of CA by many researchers in different fields, this technique has been found in many statistical software packages, such as SPSS, Minitab, Stata, SAS, Statistica and XLSTAT, and R (Greenacre, 2007).

The advantages of correspondence analysis are that: It can show the relationship between row and column categories and inform of graphs obviously with easy calculations. It is a very effective method in visualization of the data by detecting its manner and classifying the most effective categories in the low dimensional matrix (Alhuzali *et al.*, 2019). The disadvantages of correspondence analysis were that: It presents graphical relationship only without statistical measures of this relationship, the number of dimensions is determined by the researcher and its susceptibility to outliers.

This paper intended to use the data visualization method of correspondence analysis to discover the relationship between structural chromosomal aberrations-infertility problems frequencies in low dimensional map to find the manner and trend of the data in Egyptian buffaloes under Egyptian conditions.

MATERIALS AND METHODS

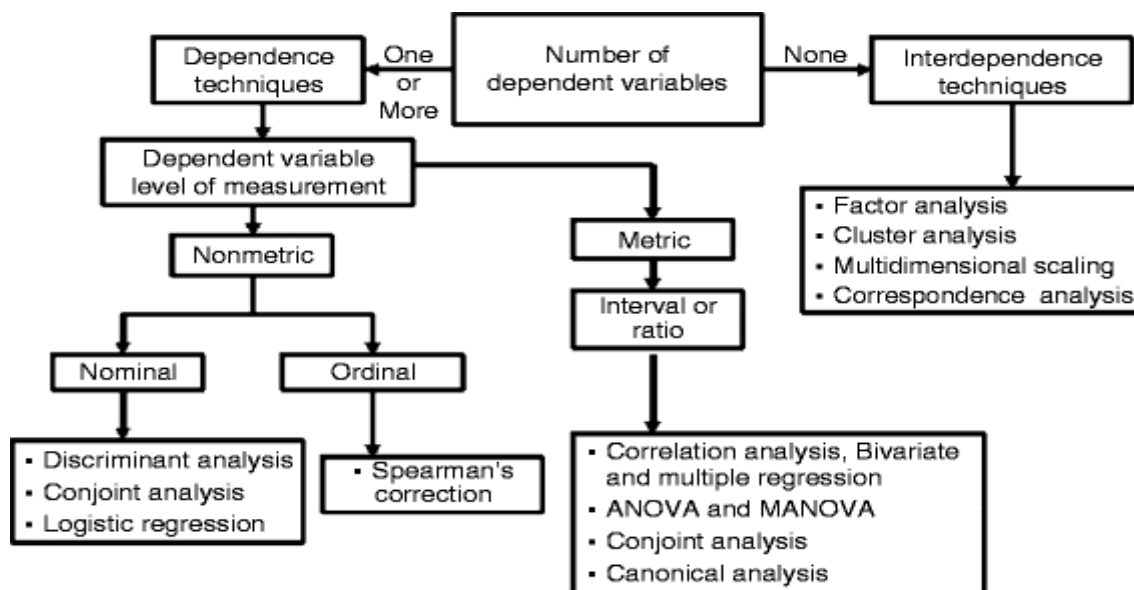
Data set:

Data were obtained from 519 animals sampled from different private farms in El-Sharkia Governorate, Egypt. Data arranged in the form of frequencies of two categorical variables (structural chromosomal aberrations and some diseases in Egyptian buffalo) arranged in a contingency table. Structural chromosomal aberrations were divided into 6 categories (gap, break, deletion, fragment, ring chromosome and centromeric attenuation). The different diseases of animals vary from normal state (without disease or control) to abnormal state. This abnormal state is related to different infertility problems (repeat breeder, anestrus, retained placenta, free-martin, vaginal prolapse, uterine prolapse, uterine torsion and habitual abortion).

The categories were arranged in a 6 x 9 contingency table representing 519 data values. Associations between variables and

among sub-categories were examined using correspondence analysis. The statistical calculations were applied by SPSS 21 statistical software.

The statistical technique applied for data analysis is chosen according to a chart of multivariate techniques as follows:



Correspondence analysis (CA):

It is a statistical method in which contingency tables changed to graphical presentations, called 'maps', and other statistics (Hoffman and Franke, 1986 and Greenacre and Blasius, 1994).

In correspondence analysis, $F = I \times J$ contingency table, where I and J are the categorical variables, and f_{ij} are the frequencies of the row category i with column category j where the total for row and column was known as marginal for them,

$$f_{i+} = \sum_{j=1}^J f_{ij}, f_{+j} = \sum_{i=1}^I f_{ij}, f_{++} = \sum_{i=1}^I \sum_{j=1}^J f_{ij}.$$

$$F(i,j) = [f_{ij}], f_{ij} \geq 0 \quad (i=1, 2, \dots, I), (j=1, 2, \dots, J).$$

Let N is the grand total and the correspondence matrix (matrix of relative frequencies) is $P_{ij} = n_{ij} / n$, then the row marginal is $\sum_{j=1}^J \sum_{i=1}^I p_{ij}$ then the i -th row marginal proportion is $p_{i.} = \sum_{j=1}^J p_{ij}$ and the j -th column marginal proportion is $p_{.j} = \sum_{i=1}^I p_{ij}$. These proportions (marginal) known as row and column probability masses.

CA is depended on the profiles calculations. A profile is considered a relative frequency

(frequencies divided by their total). The profile reflects the way the categories change of both variables.

The key concepts and methodological foundations of Correspondence Analysis (CA) are thoroughly presented in the works of Clausen (1998) and Greenacre (2007, 2010), which serve as essential references for understanding and applying CA techniques.

- Marginal frequency:

It is defined as the total frequency of each row or column and can be used for calculating row and column profiles and masses.

- Grand total:

N which is known as the total margin frequency of all rows or all columns.

- Row and column profile:

It is obtained by dividing the frequency of each element by the margin frequency of that row or column.

- Row and column mass:

It is obtained by dividing the margin frequency of column/row by the grand total frequency.

- Centroid:

It is the geometric center of the graph. The row and column profiles and mass have a major effect on the centroid of the graph.

- The total inertia:

The total inertia is the second part of model evaluation (Chi square value is the first evaluation). It explains the level of associations between rows and columns and equals variance. It is the dispersion of data from the centroid, which is the origin of the coordinate system (Diederich *et al.*, 1961). The value of the total inertia ranged from 0 to 1, high value indicates good model fit and a higher spread of data from the centroid or high variance.

The threshold of total inertia value is not documented, so its low value is taken into consideration.

The total inertia was decomposed to define significant information, which is important in describing this relationship. Many decompositions give different interpretations of the relationship, and different graphical outputs. The commonly used type of decomposition is singular value decomposition (SVD).

-Number of dimensions:

It is known that n-1 categories of the variable with the highest number of categories in the data under study explain 100% of the total inertia. In this study, groups of animals had the highest number of categories (9 categories), so the dimensions were 8-D visualization. The role of CA is dependent on decreasing the number of dimensions displayed on the visualization with the maintenance of all interpretable information.

- Chi-square value (a weighted Euclidean distance):

This is a measure of the difference between the observed and expected frequencies. It used in CA to measure the relation between categories of each variable.

The strength of association between the rows scores and column scores is measured as follows:

For complete independence of the model between the rows and columns

$$p_{ij} = p_{i.}p_{.j}$$

For difficulty of achieving complete independence, then a multiplicative measure is

$$p_{ij} = \alpha_{ij}p_{i.}p_{.j}$$

In case of complete independence of the model then, $\alpha_{ij} = \frac{p_{ij}}{p_{i.}p_{.j}} = 1$ for all $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$.

Then the Pearson chi-squared statistic $\chi^2 = n \sum_{i=1}^I \sum_{j=1}^J p_{i.}p_{.j}(\alpha_{ij} - 1)^2$.

In CA χ^2/n is the total inertia of the contingency table for showing the relationship between two categorical variables.

The total inertia is decomposed for detecting important sources of information for explaining this association. In CA, singular value decomposition (SVD) is the most important decomposition.

-Singular value decomposition (SVD):

SVD was defined by (Eckart and Young, 1963) as the square roots of the eigenvalues of a square, symmetric matrix calculated from the differences between observed and expected profiles. It is the most popular method used to decompose the Pearson ratios. Eckart & Young (1936) showed that the Pearson ratio may be decomposed into components by

$$\alpha_{ij} = \sum_{m=0}^{M^*} a_{im}\lambda_m b_{jm}$$

where $M^* = \max(I, J)$ is the maximum number of dimensions required to graphically depict the association between the row and column responses.

The relationship between the variables (diseases and aberrations) is firstly examined

by Chi square of independence for significance, but the nature of the relationship is unknown (Alhuzali *et al.*, 2019). CA assess this relationship visually. It calculates (SVD) which is considered the eigenvalue decomposition (EVD). The eigen values and eigenvectors are calculated (Hawkins, 1982 and Friendly, 1994).

The Table of frequencies becomes distances between single rows and columns in a low dimensional space. Then, a low-dimensional joint map of objects and categories in Euclidean space R_p is formed. Chi square distances of rows and columns are found. Chi square distances of rows calculated on the profiles of the rows of a matrix, where the row profile i is the vector of conditional proportions f_{ij}/f_{i+} . The Chi square distances between rows i and i' of table F is given by

$$d^2(i, i') = n \sum_{j=1}^J \frac{\left(\frac{f_{ij}}{f_{i+}} - \frac{f_{i'j}}{f_{i'+}} \right)^2}{f_{+j}}$$

The amount of the difference between the profiles of rows i and i' is $d^2(i, i')$.

Assumptions of Correspondence analysis:

There are four assumptions (Doey and Kurta, 2011).

1. Data must be categorical, not interval or ratio with independence of the variable.
2. The category of each variable is not less than 3 (more than 3).

3. Achieving homogeneity of variance between variables (variable has a value of zero for all its entries, this assumption is violated).

4. Non negative data.

For good interpretation of CA results, four steps were taken into consideration. First, the overall model should be evaluated for fitting the data, then selecting the dimension numbers, followed by measuring the contributions of row and column, and the last one was the evaluation of the visualization.

RESULTS

The overall model depended on Chi square independence test. The value of the test was 92.234, which was highly significant ($P < 0.01$) as in Table (1). This indicated that both variables under study were associated. Other statistics, such as total inertia, which was 0.178 explained the variance, the number of dimensions for interpreting and visualization was detected. Singular value also was calculated.

In summary, Table (1) shows that chromosomal aberrations and groups of animals were significantly related, and a 4-Dimension explanation is suitable. As shown in Table (1), dimension 1 could interpret dispersion in data by 53.5%, while dimension 2 could interpret dispersion in data by 27.2%. The closeness of categories in each Figure showed the similarity of them.

Table 1: Correspondence table summary.

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	0.308	0.095			0.535	0.535	0.039	0.006
2	0.220	0.048			0.272	0.807	0.044	
3	0.140	0.020			0.111	0.918		
4	0.120	0.014			0.081	0.999		
5	0.012	0.000			0.001	1.000		
Total		0.178	92.234	0.000 **	1.000	1.000		

The row profile results (disease status) were in Table 2 which gave information on graphing the row points in the final biplot. The 'mass' column indicated the percent of each disease group in relation to all other disease groups. The column 'score in dimension' indicated the coordinates in

dimension (1 and 2), where each row category will fall on the biplot. Inertia reflected the variance. The 'contribution' column reflected how each of the points loaded onto each of the dimensions, and how well the extraction of dimensions explains each of the points.

Table 2: Contributions of the row profile (disease status).

Disease status	Mass	Score in Dimension		Inertia	Contribution				
		1	2		Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		Total
					1	2	1	2	
Control	0.094	-0.406	0.553	0.014	0.050	0.131	0.354	0.468	0.822
Repeat breeder	0.146	-0.328	0.327	0.012	0.051	0.071	0.414	0.293	0.706
Anestrus	0.110	-0.521	0.322	0.015	0.097	0.052	0.604	0.165	0.768
Freemartin	0.064	0.375	-0.605	0.009	0.029	0.106	0.303	0.561	0.865
Retained placenta	0.127	0.329	-0.241	0.015	0.045	0.034	0.276	0.105	0.381
Vaginal prolapse	0.148	0.044	-0.307	0.007	0.001	0.064	0.013	0.458	0.472
Uterine prolapse	0.131	0.164	-0.694	0.016	0.011	0.287	0.070	0.892	0.962
Abortion	0.083	-0.880	-0.042	0.028	0.208	0.001	0.707	0.001	0.708
Uterine torsion	0.096	1.275	0.762	0.062	0.508	0.255	0.773	0.197	0.970
Active Total	1.000			0.178	1.000	1.000			

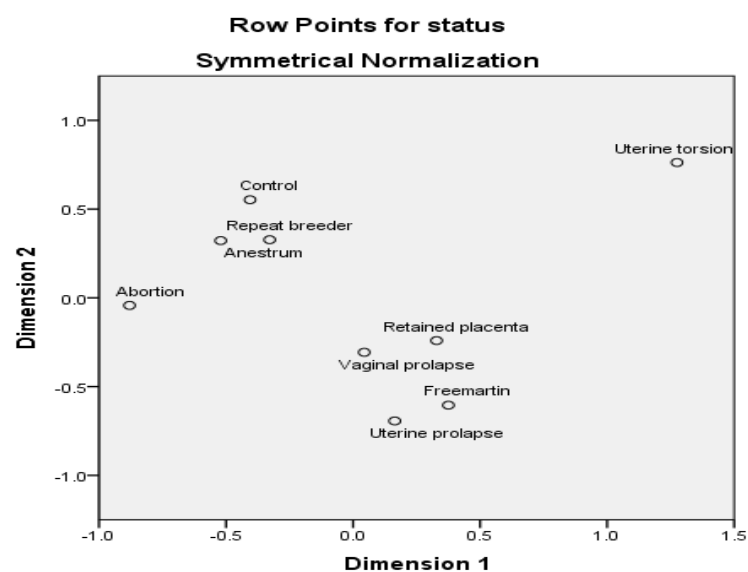


Figure 1: Row profiles plotted on a biplot (correspondence analysis for the first variable disease groups).

Column profile results (the variable structural chromosomal aberrations) were shown in Table 3, which gave information on graphing the column points in the final biplot. The 'mass' column indicated the

percent of each structural aberration in relation to all other structural aberrations. This table is similar to the above row profile table in its compositions.

Table 3: Contributions of the column profile (divisions of structural chromosomal aberrations).

Divisions of structural aberrations	Mass	Score in Dimension		Inertia	Contribution				
		1	2		Of Point to Inertia of Dimension	Of Dimension to Inertia of Point			
						1	2	Total	
Gap	0.137	-0.028	0.258	0.014	0.000	0.042	0.002	0.143	.146
Break	0.249	-0.457	-0.144	0.025	0.168	0.023	0.636	0.045	0.681
Deletion	0.272	-0.345	-0.130	0.019	0.105	0.021	0.512	0.052	0.564
Fragment	0.096	0.643	1.219	0.046	0.129	0.651	0.269	0.687	0.956
Ring chromosome	0.037	-0.877	0.513	0.014	0.091	0.044	0.599	0.146	0.745
Centromeric Attenuation	0.210	0.862	-0.478	0.059	0.506	0.219	0.819	0.179	0.998
Active Total	1.000			0.178	1.000	1.000			

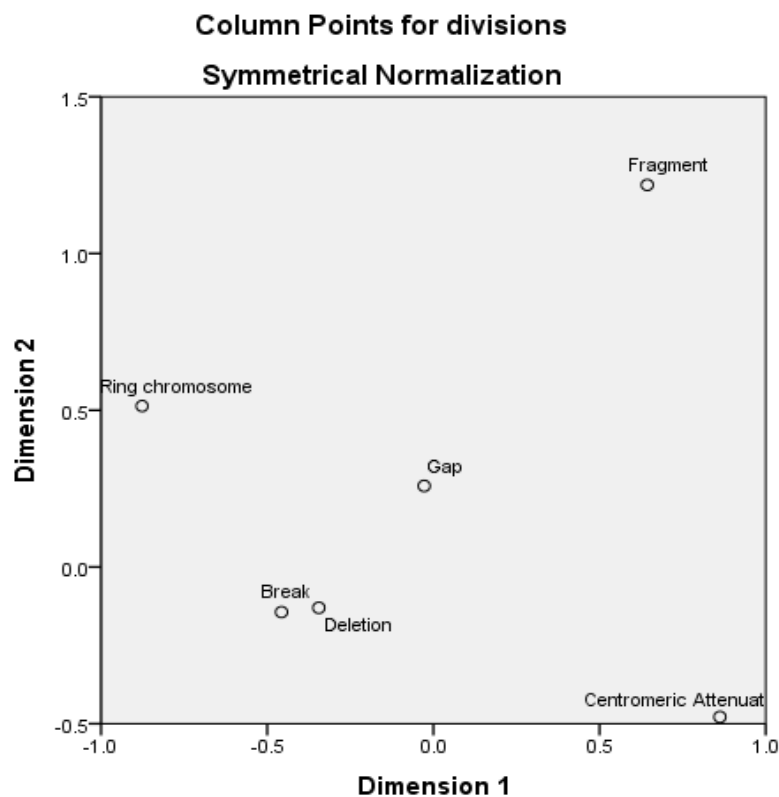


Figure 2: Column profiles plotted on a biplot (correspondence analysis for the second variable divisions of structural chromosomal aberrations).

Dispersion in data explained through dimension 1 and dimension 2, which was 53.5%, and 27.2%, respectively as a result of the row and the column profile together

(Figure 3). The closeness of categories showed the similarity between them, so the closeness of row and column between one another was examined.

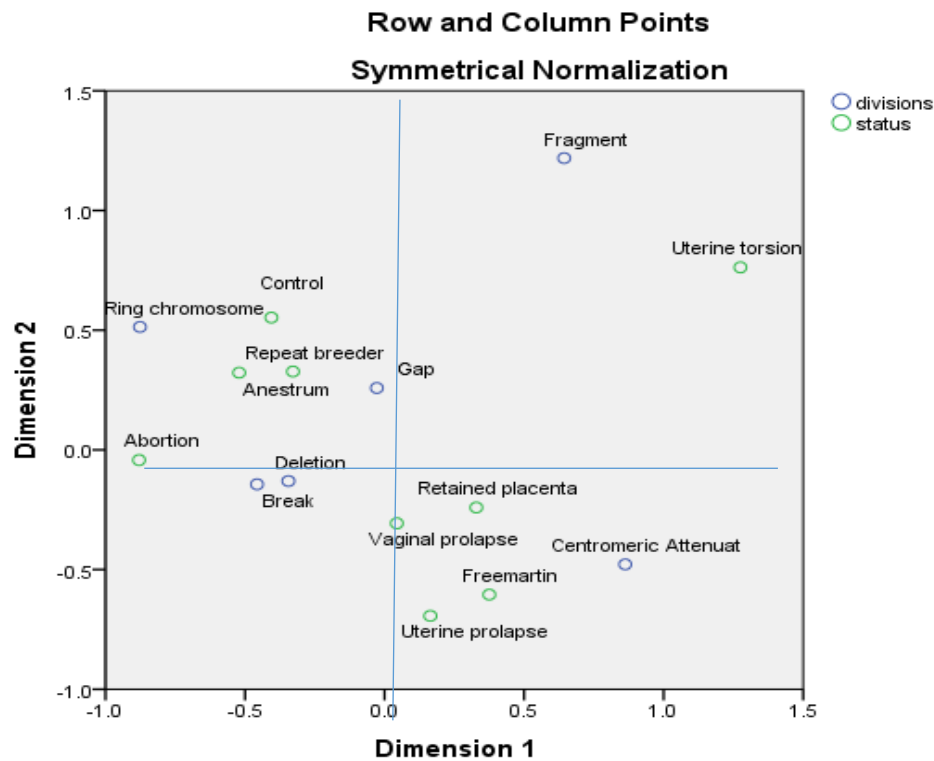


Figure 3: Biplot explaining correspondence analysis for variables, disease groups and divisions of structural chromosomal aberrations.

DISCUSSION

The study's main aim is to assess the relationship between chromosomal aberrations and infertility groups in buffalo. Our findings based on Chi-square test reveal a significant association exists between chromosomal aberrations and infertility, confirming a relationship between these variables. However, the test used does not identify which specific categories are linked or the strength of these associations (Lam, 2016). This limitation highlights the need for complementary methods, such as correspondence analysis, to gain a more detailed understanding of the relationships among categorical variables. Although MCA is less commonly applied than Chi-square or logistic regression, its ability to represent relationship patterns graphically enhances interpretability

and provides richer insights into the data structure (El-Araby *et al.*, 2023).

Correspondence analysis is increasingly used nowadays in the veterinary field. El-Araby *et al.* (2023), applied correspondence analysis to assess the relationship between categorical variables related to bluetongue disease in small ruminants. Other researchers used a combination of MCA and hierarchical cluster in studying novel goose parvo virus disease pathogeny (Lebdah *et al.*, 2025b), and locomotor disorder (Lebdah *et al.*, 2025a).

The total inertia level (total explained variance) is 0.178. This value indicated that chromosomal aberrations and disease groups in the CA explained 17.8% of the total inertia, and both variables explained each other. This value was not relatively high, but Chi square

value was still highly significant, as stated before (Doey and Kurta, 2011).

Explained variance amount arrange each dimension in the model. The major part of variance in the model was interpreted by dimension 1, followed by dimension 2 and so on. Dimension 1 interpreted 9.5% of the total 17.8% of variance accounted for. Furthermore, Dimension 2 interpreted 4.8% of the total 17.8% of variance accounted for. Dimension 3 interpreted 2% of the 17.8% of variance accounted for. Dimension 4 interpreted 1.4% of the 17.8% of variance accounted for. Dimension 5 interpreted 0% of the total variance, so it removed.

Form Table 1, the values in the proportion of inertia column gave the percent of dispersion that each dimension interpreted of the total variance explained by the model. Dimension 1 accounted for 0.535 (53.5%) of the total inertia, dimension 2 accounted for 0.272 (27.2%) of the total inertia, and dimension 3 accounted for 0.111 (11.1%) of the total inertia. Dimension 4 accounted for 0.081 (8.1%) of the total inertia, dimension 5 accounted for 0.001 (0.01%) of the total inertia, since dimensions 1, 2, 3 and 4 accounted for a large part of the total inertia (91.8%), so a 4-D performance was significantly easier to explain than a 5-D one. The rule is that more dimensions were added only if the total inertia for the first two dimensions is less than 75%, these results agreed with (Glynn and Robinson, 2014).

The singular values (the root of the eigenvalue) were 0.308, 0.220, 0.140, 0.120 and 0.012 for the first, the second, the third, the fourth and the fifth dimensions respectively. These values were arranged in an ascending manner, depending on the percentage of the variance in the column proportion of inertia. Four dimensions (primary axis) gave explanation quality by 99.9% (53.5% + 27.2% + 11.1% + 8.1%), which suggested that only 4 dimensions were used to interpret data variance completely by 99.9% as in Table 1.

After selecting dimension numbers, variables contributing to the total inertia were examined. Firstly, contribution threshold is detected. It is obtained by dividing 100 by the total number of categories in a row or column. The row variable had nine categories, the contribution threshold was 11% ($100/9 = 11.11$).

The contribution threshold was detected, then each row point contribution was examined for each dimension separately: Dimension 1 (on the x-axis) and Dimension 2 (on the y-axis). As in Table 2, uterine torsion (50.8%) and abortion (20.8%). This indicated that both disease conditions provided were highly contributed to interpret data variance in dimension 1 while other diseases were low contributed.

For dimension 2, points for uterine prolapse (28.7%), uterine torsion (25.5%), and control (13.1%) provided a high contribution to interpret dispersion in dimension 2 (the second primary axis). while other diseases were low contributed.

The relative contribution of each category (disease groups) was examined to detect the performance quality of each category in the primary axis. The contribution value is near to 1 indicating good performance quality of the category on the primary axis.

Table 2, also showed the values of relative contribution from disease groups for dimension 1, uterine torsion (0.773), and abortion (0.707) were close to 1. It indicated that the performance quality of disease groups (uterine torsion and abortion) in dimension 1 was perfect, while the performance quality of the other disease groups is less fit in dimension 1.

For dimension 2, the relative value produced by disease groups (uterine prolapse) was equal to (0.892), which indicated that uterine prolapse was the only group with perfect performance quality in dimension 2, while the other groups have less fit performance quality in dimension 2. After obtaining

coordinates for each point, the next step is to place those coordinates in the plot as shown in Figure (1).

As shown in Figure 1, uterine torsion (50.8%) and abortion (20.8%) both contributed more than 11.11% to dimension 1 and more deviated from the center than other groups. On dimension 2, uterine prolapse (28.7%), uterine torsion (25.5%), and control (13.1%) contributed significantly more than 11.11% and deviated from the center more than other groups.

The column categories were 6 (six categories in structural chromosomal aberrations). So, a contribution threshold for column points was 16.66% (100/6). The column contribution examined. On dimension 1, centromeric attenuation (50.6%) contributed significantly more than the column contribution threshold of 16.66%. So, it interpreted data dispersion in dimension 1 (the first primary axis) and other categories gave small contributions. On dimension 2, fragment (65.1%), and centromeric attenuation (21.9 %) contributed more than 16.66% in explaining data dispersion for each dimension.

Relative contribution for showing performance quality of each category in the primary axes. From Table (3), it shows that for dimension 1, the value of relative contribution for centromeric attenuation (0.819) was near to 1. It indicated that the performance quality of centromeric attenuation in dimension 1 was fit while other structural chromosomal aberrations categories were less fit representation quality in dimension 1.

In dimension 2, the relative value of all categories was far from 1, which indicated that the performance quality of structural chromosomal aberrations for all categories in dimension 2 was not fit well.

In Figure (2), centromeric attenuation (50.6%) contributed more than 16.66% to dimension 1 and more deviated from the center than other groups. On dimension 2, fragments (65.1%), and centromeric

attenuation (21.9%) contributed significantly more than 16.66% and deviated from the center more than other groups.

All the points of rows and columns appeared in Figure (3), regardless of whether they contributed to total inertia or not. Therefore, row and column points might be clustered on a biplot, but this clustering could be the result of chance, which is similar to a type 1 error in hypothesis testing, which exists as a “false positive.

The clusters or closeness for visualizing how the disease groups and divisions of structural chromosomal aberrations categories are related. For example, according to the first dimension, fragments of structural chromosomal aberrations and uterine torsion of disease groups clustered on the upper right side of the map. Both considered highly and negatively related with other categories. Similarly, when second dimension categories located on the upper side of the map were considered negatively associated with others on the lower side, and so on (Çak *et al.*, 2023).

CONCLUSION

CA is a descriptive statistical technique that explains contingency tables and data pattern trends in a visualization manner. It aims to classify row and column relationships to find a low dimensional matrix, which can be divided into minimal parts, to be easy to analyze and explain each part. The simple correspondence method applied for two variables with ($i \times j$) contingency table, to measure the distance between column row profiles, and display it graphically. Graphical output informs of a biplot, in which the proximity of items (qualitative variable values) indicates the relationship degree between categories. CA can help with veterinary aspects easily, rather than other complex techniques. In short, it can be applied for dimension reduction and data simplification, clustering, classification, feature selection, knowledge extraction, visualization and pattern discovery.

REFERENCES

- Abdi, H. and Valentin, D. (2007):* "Multiple correspondence analysis," in Encyclopedia of Measurement and Statistics, N. Salkind, Thousand Oaks, CA, USA: Sage, pp. 1–13.
- Agresti, A. (1996):* An Introduction to categorical data analysis, John Wiley & Sons, New York, NY, USA. (1st ed).
- Alhuzali, T.; Stojanovski, E. and Beh, E.J. (2019):* Correspondence analysis approach to examine the Nobel Prize. 23rd International Congress on Modelling and Simulation, Canberra, ACT, Australia.
- Beh, E.J. (2007):* "Simple correspondence analysis of nominal-ordinal contingency tables. Journal of Applied Mathematics, 2008(1).
- Benzecri, J.P. (1992):* Correspondence Analysis Handbook, New York: Marcel Decker.
- Bond, J. and Michailidis, G. (1997):* "Interactive correspondence analysis in a dynamic object-oriented environment. Journal of Statistical Software, 2(8).
- Çak, B.; Keskin, S.; Yilmaz, O. and Demirel, A.F. (2023):* Correspondence analysis to visualize the relationships between alpha-S1 casein and beta-lactoglobulin gene polymorphisms in norduz sheep. Van Vet J, 34 (2) 163-167.
- Clausen, S.E. (1998):* Applied correspondences analysis: An introduction, Thousand Oaks, CA: Sage.
- De Leeuw, J. (1983):* On the prehistory of correspondence analysis. Statistica Neerlandica, 37: 161–164.
- Diederich, P.; French, J. and Carlton, S. (1961):* "Factors in judgments of writing ability," ETS Res. Bull. Ser., no. 2, pp. i–iv, 1–93.
- Doey, L. and Kurta, J. (2011):* Correspondence analysis applied to psychological research. Tutorials in Quantitative Methods for Psychology, Vol. 7(1), p. 5-14.
- Eckart, C. and Young, G. (1936):* The approximation of one matrix by another of lower rank. Psychometrika, 1, 211–218.
- El-Araby, I. E., Moawed, S. A., Hassan, F. A., and Gouda, H.F. (2023):* A combined approach of multiple correspondence analysis and hierarchical cluster analysis for profiling relationships among categorical risk predictors: a bluetongue case study. Slovenian Veterinary Research, 60.
- Friendly, M. (1994):* Mosaic displays for multi-way contingency tables. Jasa, 89:190-200.
- Glynn, D. and Robinson, J.A. (2014):* "Correspondence analysis: Exploring data and identifying patterns," in corpus methods for semantics: Quantitative studies in polysemy and synonymy, Philadelphia, PA, USA: John Benjamins, pp. 443–486.
- Greenacre, M. and Blasius J. (1994).* Correspondence analysis in the social sciences, London; Academic Press, (1st ed.).
- Greenacre, M.J. (2007):* Correspondence analysis in practice, London: Chapman & Hall / CRC Press, (2nd ed.).
- Greenacre, M.J. (2010):* Biplots in practice. Spain: Fundacion BBVA. Available at: <http://www.multivariatestatistics.org/biplots.html>.
- Hawkins, D.M. (1982):* Topics in applied multivariate analysis, USA: Cambridge University Press, 183-206 (1st edition).
- Hill, M.O. (1974):* Correspondence analysis: a neglected multivariate method. Applied Statistics, 23:340-354.
- Hoffman D.C. and Franke, G.R. (1986):* Correspondence analysis: graphical representation of categorical data in marketing research. Journal of Marketing Research, 23:213-27.
- Kumar, N. and Nasser, M. (2016):* Multivariate data visualization: Correspondence Analysis, Classical and Robust Singular Value Decomposition and depth based approach. International Journal of Advanced Research, 4 (1): 416- 425.

- Lam, C. (2016): Correspondence analysis: A statistical technique ripe for technical and professional communication researchers. IEEE Transactions on Professional Communication · VOL. 59, NO. 3.
- Lebdah, M.A.; Eid, A.A.M.; ElBakrey, R.M.; El-Gohary, A.E.; Mousa, M. R., Gouda, H.F.; Gad, A.F.; Helal, S.S. and Seadawy, M.G. (2025a): Novel goose parvovirus in naturally infected ducks suffering from locomotor disorders: molecular detection, histopathological examination, immunohistochemical signals, and full genome sequencing. Avian pathology: journal of the W.V.P.A, 54(2), 241–254. <https://doi.org/10.1080/03079457.2024.2419038>.
- Lebdah, M.A.; Eid, A.A.M.; ElBakrey, R.M.; El-Gohary, A.E.; Seadawy, M.G.; Mousa, M.R.; Gouda, H.F.; Goda, N.I.A.; El-Hosseney, M.F.; El-Tahlawy, A.S.; Sami, R.; Al-Eisa, R.A. and Helal, S.S. (2025b): Egyptian Novel Goose Parvovirus in Immune Organs of Naturally Infected Ducks: Next-Generation Sequencing, Immunohistochemical Signals, and Comparative Analysis of Pathological Changes Using Multiple Correspondence and Hierarchical Clustering Approach. Viruses, 17(1), 96. <https://doi.org/10.3390/v17010096>.
- Pearson, K. (1906): On certain points connected with scale order in the case of a correlation of two characters for some arrangement give a linear regression line. Biometrika, 5:176–178.
- SPSS V21. IBM Corp. Released (2017): IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY:IBM Corp.

تحليل التناظر لتصنيف العلاقات بين فئات المتغيرات النوعية (الانحرافات الكروموسومية التركيبية - العقم في الجاموس المصري)

فاطمة دسوقي محمد عبدالله ، أسماء وجيه زغلول

E-mail: moibho_stat2021@yahoo.com; asmaa.wagih2008@gmail.com

Assiut University web-site: www.aun.edu.eg

تهدف هذه الدراسة إلى تحليل البيانات العددية باستخدام التحليل الإحصائي المعروف بتحليل التناظر. يطبق تحليل التناظر لتحديد طبيعة العلاقة بين فئات المتغيرات النوعية في الجداول المقارنة. حيث أن المتغيرات قيد الدراسة كانت عبارة عن الانحرافات الكروموسومية التركيبية وبعض الأمراض المتعلقة بالعقم في الجاموس المصري ، حيث كانت الانحرافات الكروموسومية التركيبية مقسمة إلى ٦ فئات عبارة عن (break و gap و ring chromosome و fragment و deletion) وقسمت المجموعات المرضية في centromeric attenuation في الحيوانات إلى ٩ فئات وهي (control و repeat breeder و anestrus و retained placenta و free-martin و habitual abortion و uterine prolapse و vaginal prolapse و uterine torsion). أظهرت النتائج أن قيمة مربع كاي لإستقلالية العوامل معنوية مما يدل على أن المتغيرين مرتبطين معاً. وكانت قيمة التباين الكلي المشروح أسبابه ١٧,٨٪ مما يدل على أن الانحرافات الكروموسومية التركيبية و مجموعات الحيوانات المرضية المتعلقة بالخصوبة يفسر كل منهما الآخر بنسبة ١٧,٨٪. بالنسبة للبعد الأول والثاني كانا أكثر تفسيراً للتباين حيث فسر البعد الأول التباين بنسبة (53.5%) والبعد الثاني فسر التباين الموجود في البيانات بنسبة (27.2%) من المجموع الكلي للتباين. abortion و uterine torsion كانا الأكثر مشاركة في تفسير التباين بالنسبة للبعد الأول ، كما أن البعد الثاني كانت الأقسام uterine prolapse و uterine torsion و control أكثر تمثيلاً له بين فئات المتغير الأول. أما بالنسبة للمتغير الثاني كانت المجموعة centromeric attenuation الأكثر معنوية بالنسبة للبعد الأول وكانت fragment و centromeric attenuation الأكثر مشاركة في تفسير التباين بالنسبة للبعد الثاني عن باقي أقسام متغير الانحرافات الكروموسومية التركيبية. يمكن لمربو الحيوانات استخدام تحليل التناظر في مزارعهم لتسهيل فهم نمط البيانات والتمثيل البياني لها في حالة القيم الكثيرة للبيانات.