YYÜ TAR BİL DERG
(YYU J AGR SCI)
2014, 24(2): 121- 126

Geliş Tarihi (Received): 21.08.2013
Kabul Tarihi (Accepted): 07.05.2014

Araştırma Makalesi/*Research Article (Original Paper)*

Directional Data Analysis and An Application

Gazel SER

University of Yuzuncu Yil, Agriculture Faculty, Department of Animal Science, Biometry and Genetic Unit, Van, Turkey Corresponding author : gazelser@yyu.edu.tr

Abstract: Directional data analysis method is used in the observed values that are obtained angularly. In this technique, every observation obtained as an angular data type is located on the unit circle axis and a mean vector in relation to the data is formed. The statistical computations are obtained through that mean vector. In this study, some of the basic directional statistical methods are aimed to be analyzed. In accordance with this purpose, the angles between teats are measured six times in the mornings before milking the goats during the lactation period. Some directional statistics were obtained from this analysis by these teat angles measurements were analyzed. As a consequence, the length of mean vector increased while the circular variance was decreasing. The lowest variance and the maximum length of the mean vector are obtained in the 6th test milking. In addition, according to the Watson-Williams F test, which is used to determine whether there is a difference between the mean angles, the differences between the mean angles of test milking 5 and 6 are found to be statistically important ($p<0.05$) and the mean angles decrease through the end of the lactation period.

Key words: Directional data analysis, circular scale, circular statistics, teat angle

Yönlü Veri Analizi ve Bir Uygulama

Özet: Yönlü veri analiz yöntemi, açısal olarak elde edilen gözlem değerlerine uygulanır. Yöntemde, açısal veri tipinde elde edilen her gözlem birim çember ekseni üzerine yerleştirilir ve verilere ait bir ortalama vektör elde edilir. İstatistiksel hesaplamalar bu ortalama vector üzerinden yapılır. Bu çalışmada, bazı temel yönlü veri analiz yöntemlerinin incelenmesi amaçlanmıştır. Bu amaç doğrultusunda keçilerin meme başları arası açıları, sabah sağımlarından önce laktasyon periyodu boyunca toplamda altı kez yapılan süt kontrollerinde alınmıştır. Meme başları arası açısal ölçümler, bu analiz yöntemiyle analiz edilerek bazı yönlü istatistikler elde edilmiştir. Sonuç olarak, ortalama vektör uzunluğu arttıkça dairesel varyans azalmıştır. En düşük varyans ve en yüksek ortalama vektör uzunluğu 6. test sağımından elde edilmiştir. Ayrıca ortalama açılar arasında farklılık olup-olmadığının belirlenmesinde uygulanan Watson-Williams F testi sonuçlarına göre, 5. ve 6. test sağımlarının ortalama açıları arasındaki farklılık istatistiksel olarak önemli ($p<0.05$) bulunmuş ve laktasyonun periyodunun sonuna doğru ortalama açılar azalmıştır.

Anahtar kelimeler: Yönlü veri analizi, dairesel ölçek, dairesel istatistikler, meme başları arası açı

Introduction

Directional data analysis is the subfield of the statistics (Berens and Velasco 2009). The observations obtained from the directional data cannot be analyzed with standard statistical methods. It is because that kind of data inherently are circularly structured and they are located on the 360 degree axis (Açar 2010). The measuring instruments used to get such data are clock, compass and protractor. Every observation obtained by this way is located on unit circle (Gaile and Burt 1980**).** The observations of different data types (like day or month) can be transformed into angles. For instance, time arrivals of the patients going to the accidents unit of a hospital in 24 hours, directions of the winds on a daily basis, directional movements of the animals, directions of the bees after a stimulus in a hive or migration directions of the birds (Mardia and Jupp 2000; NCSS 2012). Traditional statistical methods may not characterize enough the statistical specifications of the data obtained from the circularly scaled variables (Mahan 1991). For example, the arithmetic mean of 10˚, 30˚ and 350˚ is 130˚ and they show south east as the geographical direction. However, when all the directions are located on circle, they show a direction towards the north. Therefore, the calculated classic arithmetic mean may lead a wrong direction. In this case, using special methods such as directional data analysis methods instead of the classic methods is more appropriate to determine the angular measures (Gaile and Burt 1980; Mahan 1991; Jones 2006; Berens and Velasco 2009). In this study, some of the basic directional data analysis methods are aimed to be analyzed.

Material and Methods

Animal Sampling

Thirty Akkeçi crossbred goats used as material in the research began to be milked after kids weaned. Udder measurements and milk controls were performed for 6 times with intervals of two weeks until the end of the lactation period. The angle between right and left teats among the measurements were used in this study. The measurements were taken in the morning and immediately before milking the animals. No further measurement was taken prior to evening milking. Teat angle were measured according to Kor *et al.*, 2004.

Statistical Analysis

Directional basic statistics

To calculate the mean angle, the cartesian coordinates ($\overline{C}, \overline{S}$) of the data points must be computed;

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n}\cos\theta_i \ , \ \overline{S} = \frac{1}{n}\sum_{i=1}^{n}\sin\theta_i$$

Mean angle ($\overline{\theta}$) is the solution of the equations,

$$\overline{C} = \overline{R}\cos\overline{\theta}_i \ , \ \overline{S} = \overline{R}\sin\overline{\theta}_i$$

These values allow for the computation of the length of the mean vector ($\overline{R}$),

$$\overline{R} = (\overline{C}^2 + \overline{S}^2)^{1/2} \ , \ 0 \le \overline{R} \le 1$$

When the data points equally distributed on the circle, length of mean vector becomes ‘’0’’. However, when the data points are on the same location, length of mean vector is ‘’1’’. The mean angle ($\overline{\theta}$) is,

$$\overline{\theta} = \begin{cases} \tan^{-1}(\overline{S}/\overline{C}) & \overline{C} > 0, \overline{S} > 0 \\ \tan^{-1}\left(\overline{S}/\overline{C}\right) + \pi & \overline{C} < 0 \\ \tan^{-1}\left(\overline{S}/\overline{C}\right) + 2\pi & \overline{C} > 0, \overline{S} < 0 \end{cases}$$

Note that in the context of circular statistics $\overline{\theta}$ does not mean ($\theta_1 + ... + \theta_n / n$) (Mardia and Jupp 2000). Circular variance is used as a measure of dispersion around the mean vector. The simplest of these is the sample *circular variance* defined as,

$$V = (1 - \overline{R}), \ 0 \le V \le 1$$

(Mahan 1991; Mardia and Jupp 2000; NCSS 2012).

Watson-Williams F-test is used to determine statistical different between two or more mean angles . Watson-Williams test is a modification of the linear F test (Mahan 1991; Mardia and Jupp 2000; Hassan *et al*. 2009).

Results and Discussion

In Table 1, circular descriptive statistics of angles between right and left teats obtained for six test milking were given.

Table 1. The results of circular descriptive statistics

Time of test milking	1	2	3	4	5	6	Total
Data Type	**Angles**	**Angles**	**Angles**	**Angles**	**Angles**	**Angles**	**Angle**
Numberof Observation	30	29	29	18	17	17	140
Mean angle ($\bar{\theta}$)	32.222˚	32.376˚	32.396˚	32.633˚	34.118˚	30.882˚	32.412˚
Length of Mean Vector($\bar{R}$)	0.995	0.997	0.996	0.998	0.998	0.998	0.997
Median	32˚	32˚	33˚	33˚	34˚	30˚	33˚
Circular Variance (V)	0.005	0.003	0.004	0.002	0.002	0.002	0.003
Circular Standart Deviation	5.872˚	4.52˚	4.973˚	3.512˚	3.411˚	3.693˚	4.69˚
Standart Error of Mean	1.072˚	0.839˚	0.908˚	0.806˚	0.827˚	0.896˚	0.394˚
%95 Confidence Interval	30.121˚ 34.324˚	30.731˚ 34.021˚	30.616˚ 34.176˚	31.053˚ 34.213˚	32.496˚ 35.74˚	29.126˚ 32.637˚	31.64˚ 33.184˚

The mean angle of all time of test milking is 32.412˚. If observation values are calculated using classical arithmetic mean, the means of all time of test milking shall be obtained as 32.564. This classic arithmetic mean is a value close to mean angle (see Table 1). However, this situation is misleading. Because the circular mean calculated for the circular data set is not the measurement of the location of data such as classic arithmetic mean but rather it is a function of the zero direction and choice of the turning direction. Therefore, descriptive statistics obtained with classic statistical methods, such as the arithmetic mean are not the appropriate methods for the circular data (Gaile and Burt 1980; Mahan 1991; Jones 2006; Peker and Bacanlı 2009; NCSS 2012). Moreover, directional data analysis unlike the linear-based analysis methods does not get affected by the extreme values (outlier) in the data (Berens and Velasco 2009). The length of the mean vector ($\bar{R}$) plays a determining role in the directional data analysis. Especially, the length of the $\bar{R}$ vector affects the estimate of variance. For example, when the length of the mean vector decreases from 1 to 0, variance of the distribution increases. In other words, when the length of the mean vector increases from 0 to 1, variance of the distribution decreases so it is thought to be a measurement of the variance dispersion (Mahan 1991). Table 1 shows that when the length of the mean vector is 0.995, the variance is 0.005 in test milking 1 when $\bar{R} = 0.998$, the variance is 0.002 in test milking 6 which means that when the length of the mean vector increases, the circular variance decreases. Nevertheless, in

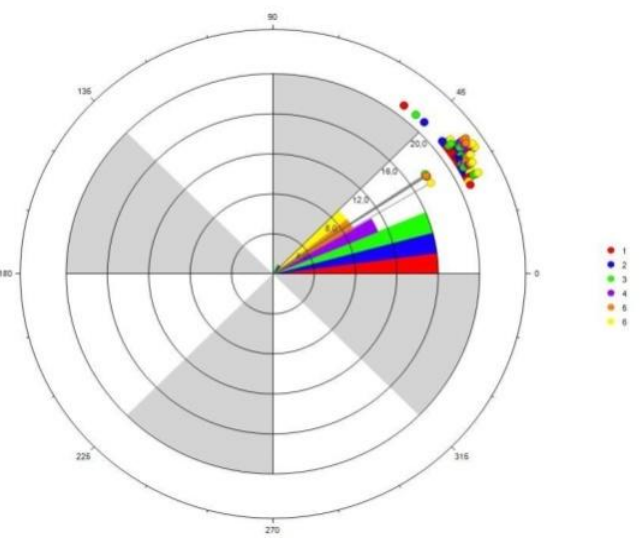

Figure 1. Circular plot of teat angle

Figure 2 (Test milking 1) is more distributed while in Figure 7 (Test milking 6) there is a uniform dispersion around the mean vector. Circular plot (also rose diagram) of distribution teat angles measured for all time of test milking is depicted in Figure 1.

On the circle, 0 degrees is on the east, 90 degrees is at the north, 180 degrees is on the west and 270 degrees is at the south. In the Figure 1, the values of angles between teats were between 0° and 45°. However since the 3 values (53°, 46° and 49°) obtained in test milking 1, 2 and 3 are higher than 45°, they were positioned after this value. In Figure 2-7, rose plots of measurement values at 6 test milking points were given.

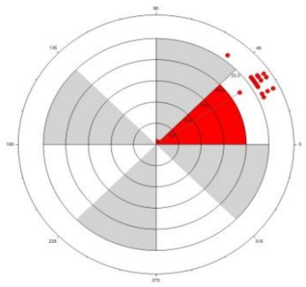

Figure 2. Rose plot of teat angle when test milking=1

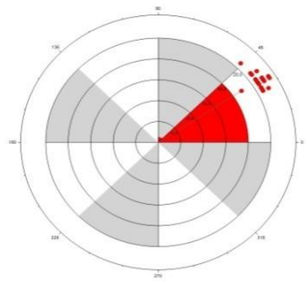

Figure 3. Rose plot of teat angle when test milking=2

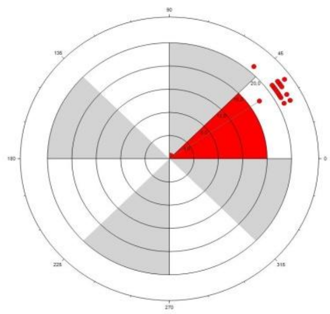

Figure 4. Rose plot of teat angle when test milking=3

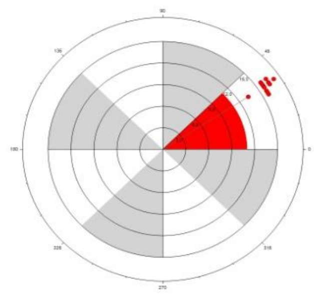

Figure 5. Rose plot of teat angle when test milking=4

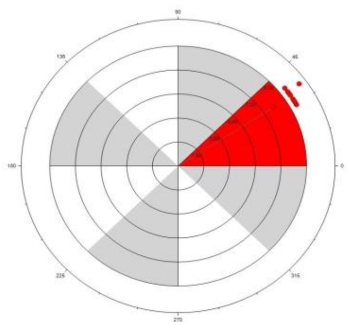

Figure 6. Rose plot of teat angle when test milking=5

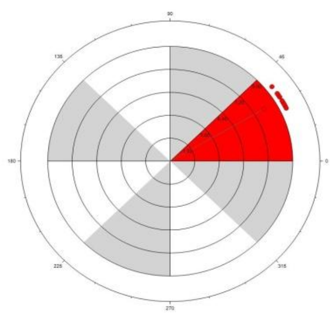

Figure 7. Rose plot of teat angle when test milking=6

The rose plots between Figures 2-7 provide an idea about the angular direction of data from the each test milking. The parallel line in these figures shows a group range and gives the number of observation within this group range. Furthermore, the long straight line apparent in each figure shows the direction of the mean regarding that test milking and gives an idea about the dispersion of observations around the mean direction (Hassan *et al.* 2009; NCSS 2012). When the length of the mean vector increases, the variation between the data decreases and it shows a uniform distribution. Therefore, the variation in Figure 7 (Test milking 6) is lesser compared to the other all time of test milking and the data shows more uniform distribution. In order to detect whether there is a difference among all time of test milking, Watson-Williams F test was applied and the results are given in Table 2.

Table 2. Watson-Williams F tests resuls in matrix format (F scores (lower half) and probabilities (upper half))

	Test milking 1	**Test milking 2**	**Test milking 3**	**Test milking 4**	**Test milking 5**	**Test milking 6**
Test milking 1	-----	0.912	0.904	0.788	0.238	0.409
Test milking 2	0.012	-----	0.987	0.838	0.185	0.264
Test milking 3	0.015	2.49E-4	-----	0.860	0.221	0.289
Test milking 4	0.073	0.042	0.032	-----	0.221	0.166
Test milking 5	1.427	1.813	1.543	1.557	-----	0.015*
Test milking 6	0.695	1.279	1.152	2.008	6.635	-----

*$p<0.05$

For our data set, the test statistics (Test milking 5 and 6) give the value of 6.635 and the probability value is 0.015. Hence, there are significant differences between the two test milking. The mean angle become smaller depends on decreasing the milk content of udder at the end of the lactation period.

Conclusion

Directional data analysis is not common method in animal science research. However, it is aimed to give some information about some of the basic statistics of the directional data analysis and their graphical representations. The measures obtained as angular differ from the classic methods because of the directions that the observed values are located on the unit circle in the directional data analysis and the processes are based on the length of the mean vector. Therefore, for the data sets that have circular scales, directional data analysis methods can be applied easily to the every field of life sciences.

Acknowledgments

I would like to thank to Professor Aşkın Kor and Assistant Professor Serhat Karaca for providing data to use in this study.

References

Açar F (2010). An application for the analysis of directional data. M. Sc. Thesis, Yüzüncü Yıl University 67p. Available from: Yüzüncü Yıl University, Institute of Natural Science.

Berens P, Velasco MJ (2009). The circular statistics toolbox for Matlab. Technical Report No. 184, Max Planck Institute for Biological Cybernetics, Germany.

Gaile GL, Burt JE (1980). Directional Statistics. http://qmrg.org.uk/files/2008/11/25-directional-statistics.pdf. (Access date: 02 Mayıs, 2014).

Hassan SF, Hussin AG, Zubairi YZ (2009). Analysis of Malaysian wind direction data using ORIANA. Modern Applied Science 3(3): 115-119.

Kor A, Keskin S, Karaca S, Arslan S (2004). The effect of age and lactation order on some udder measurments in Akkeçi. Yüzüncü Yıl University Journal of Agricultural Sciences 14(2): 105-111.

Mahan RP (1991). Circular Statistical Methods: Applications in Spatial and Temporal Performance Analysis. http://www.dtic.mil/dtic/tr/fulltext/u2/a240751.pdf. (Access date: 24 Mart, 2014).

Mardia KV, Jupp PE (2000). Directional Statistics, John Wiley & Sons, Ltd., Chichester.

Jones TA (2006). MATLAB functions to analyze directional (azimuthal) data—I: single-sample inference. Computers & Geosciences 32: 166–175.

NCSS (2012). Chapter 230: Circular Data Analysis. http://ncss.wpengine.netdnacdn.com /wpconten/themes/ncss/pdf/Procedures/NCSS/Circular_Data_Analysis.pdf. (Access date: 24 Mart, 2014).

Peker K Ö, Bacanlı S (2009). Application of the sequential test of the mean direction for the circular data when the concentration parameter is unknown. Journal of Statisticians: Statistics and Actuarial Sciences 2 (1): 9-18.