http://www.medscidiscovery.com

OPEN ACCESS JOURNAL ISSN: 2148-6832

MSD

Medical Science and Discovery
2019; 6(3):63-72

Research Article Doi: 10.17546/msd.

Application of circular statistics to life science

Yıldırım Demir[1]*, Ömer Cevdet Bilgin[2]

Abstract

Objective: The aim of the study was to explain circular statistics and hypothesis tests with birth data. The accuracy of the statistical method used in scientific research is related to the data structure and scale type. Therefore, scale types and data structures should be well defined. It is possible to frequently come across circular data in many different scientific fields such as medicine, biology and physics. These data are usually obtained by compass or clock. Compass; the flight direction of any animal that is released, the direction of the wind or the direction of current in the ocean, clock; the birth time of infants, the time of crisis, circadian rhythms or biological rhythms can be shown as examples. Apart from the clock, such data may also be obtained by a scale that expresses a time such as day, month and year.

Material and Methods: The data related to 179 normal deliveries that took place in Yüzüncü Yıl University Medical Faculty Hospital in 2008 were used. Circular data analysis was performed using the NCSS2007 statistical package program.

Results: The times of birth of infants show a uniform distribution. No significant difference at a significance level of 5% was found between the times of birth according to gender.

Conclusion: It has been stated that circular data cannot be analyzed by the analysis methods developed for linear data due to several reasons. If circular data are analyzed by linear statistical methods, inaccurate or nonsense results usually emerge. Therefore, it was emphasized that appropriate statistical methods should be used.

Key words: Circular data analysis, circular statistics, statistics, von Mises

Introduction

The suitability of statistical methods used in scientific research is directly related to the data structure and scale type. Therefore, scale types and data structures should be well defined and known by the researchers. The angular scale, which includes circular data, is usually defined within the interval scale. It is possible to frequently come across circular data related to angular scale in many different scientific fields such as medicine, biology, geology and physics (1). These data are usually obtained by compass or clock. The direction of movement of any animal that is released, the direction of the wind or the direction of current in the ocean can be shown as examples for the data obtained by compass, while the birth time of infants, the time of crisis, circadian rhythms or biological rhythms can be shown as examples for the data obtained by clock. Apart from the clock, such data may also be obtained by a scale that expresses a time such as day, month and year (2). Circular data are shown as the points on the perimeter of unit circle with central origin or the unit vectors that combine these points with origin depending on the appropriately selected zero direction and the selection of the direction of movement.

Zero direction refers to the starting point, and the direction of movement refers to clockwise or counterclockwise. Since circular data are shown on the unit circle, these observations correspond to any θo angle between 0o and 360o. θo angle is the angle between the unit vector and the starting point according to the reference direction (3).

Furthermore, the absence of a natural ordering of observations in circular data, overlapping of start and end points ($0=2\pi$), and the fact that the value θ is periodical with $\theta+p(2\pi)$ value for any p integer significantly differentiate circular data analysis from univariate and multivariate linear statistical analysis. Even though the need for measurements related to the selection of arbitrary zero direction and direction of movement does not make most of the classical linear statistical techniques and measurements nonsense, these techniques and measurements give incorrect results in many cases. In this context, the selection of appropriate statistical methods for circular data is important. In this study, circular data analysis methods were applied to the data on the times of birth of infants, and the interpretation of results and the suitability of these methods were discussed (4).

Received 03-01-2019 **Accepted** 30-03-2019 **Available Online** 30-03-2019 **Published** 30-03-2019
1 Yuzuncu Yil University, Faculty of Economics and Administrative Sciences, Dept. of Statistics, Van, TR
2 Ataturk University, Faculty of Agriculture, Dept. of Biometry and Genetics, Erzurum, TR
*** Corresponding Author:** Yıldırım Demir **E-mail:** ydemir@yyu.edu.tr **Phone:** +90 (505) 452 3329

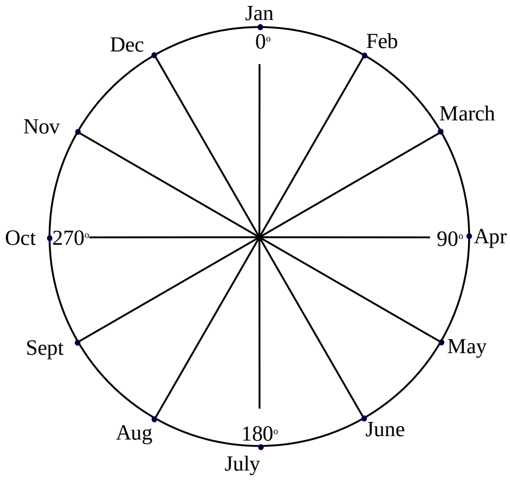

Figure 1: Circular representation of the months

Material and Methods

The research material consisted of the data related to 179 normal singleton births that took place in Van Yüzüncü Yıl University Faculty of Medicine Obstetrics and Gynecology Department between 16/04/2008 and 31/05/2008. The data set was recorded as the time of birth and baby's gender. 104 and 75 of 179 births were male and female, respectively. While all data were discussed for a single sample while evaluating the times of birth, the baby's gender (1: Male, 2: Female) was taken into account for two samples. For circular data analysis, the starting direction and the direction of rotation were taken as north and clockwise, respectively.

Whether there was any difference between the times of birth during birth events and whether there was any difference between the times of birth during birth events according to gender were evaluated by the circular data analysis method, and NCSS 2007 statistical package program was used for the evaluations (5).

Circular Descriptive Statistics: The graphical representation for getting an idea about circular data, and for the analysis and interpretation of data is one of the important stages. Different graphics and histograms are used in the representation of circular data. These graphics are used to summarize the data set and to get an idea about the data distribution before the statistical calculations.

While any point p_i in the plane is shown as (x,y) according to cartesian coordinates, it is shown as (r, θ) according to polar coordinates. Since the relationship between the directions is examined in the analysis of circular data, the vectoral sizes of data points are of no importance, and it is assumed that these data points are distributed on the perimeter of the unit circle due to the ease of operations. Thus, the distance of any point p_i to origin is 1.

Conversions are frequently performed between coordinate systems in the circular data analysis. These conversions are performed using *sine* and *cosine* trigonometric functions as the following (3).

$$cos\theta = \frac{x}{r} \ , \qquad sin\theta = \frac{y}{r} \tag{1}$$

Furthermore, the data in the original scale is converted by equation (2). Here; θ gives the angular value of the data, *a* gives the data on the original scale and *k* gives the entire cycle on the scale where *a* is measured.

$$\theta = \frac{360.a}{k} \tag{2}$$

Mean Direction: In the calculation of mean direction for a data set indicating concentration towards a direction, firstly, the mean components of the unit vectors p_i (*i*= 1, 2, ..., *n*) relative to the vertical coordinate system on the perimeter of unit circle corresponding to the angle θ_i are taken as the following,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} cos\theta_i \ , \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} sin\theta_i \tag{3}$$

According to these $\bar{x}$ and $\bar{y}$ mean components found, the mean resultant vector length is calculated as the following

$$\bar{R} = \sqrt{\bar{x}^2 + \bar{y}^2} \tag{4}$$

and according to this equation, the mean direction is calculated by means of any of the following equations

$$\bar{\theta} = cos^{-1}\left(\frac{\bar{x}}{\bar{R}}\right), \qquad \bar{\theta} = sin^{-1}\left(\frac{\bar{y}}{\bar{R}}\right) \tag{5}$$

If it is R=0, the mean direction is undefined. In this case, it is stated that the data set is not concentrated in any direction on the unit circle and does not have any mean direction. If it is R=n, it is stated that the data set has a mean direction and that all observations are concentrated in the mean direction (6, 7).

Circular Variance: Since resultant vector length (R) is a measure of scattering which shows to what extent the observations are concentrated around the center, there is a close relationship between the variance, which is a measure of dispersion around the mean, and the resultant vector length in circular data (8).

A sample scattering related to random unit vectors is expressed by

$$D = n - \sum_{i=1}^{n} cos(\theta_i - \bar{\theta}) \tag{6}$$

Accordingly, if the dispersion around the mean direction is indicated by V, so it is

$$V = \sum_{i=1}^{n} cos(\theta_i - \bar{\theta}) \tag{7}$$

and if equation (7) is resolved, the sample variance for circular data is calculated as the following

$$V = \frac{1}{n} \cdot D$$

$$V = 1 - \bar{R} \tag{8}$$

In circular data, as in linear data, as the sample variance gets smaller, the distribution becomes homogenous. However, unlike linear variance, circular variance takes values between 0 and 1. If all observations are in the same direction, in other words, if there is no scattering, the mean resultant length will be close to 1, and accordingly, the variance will be minimum. If the observations are uniformly distributed on the perimeter of the circle, in other words, if scattering is maximum, then the mean resultant vector length will be 0, and accordingly, the variance will be 1 (9, 10).

Circular Standard Deviation: The sample standard deviation for directional data is calculated by the transformation of sample variance, similar to the standard deviation on the line (4). The sample standard deviation appropriate to the circular sample variance in the range (0,1) is defined as the following

$$v = \frac{180^o}{\pi}\sqrt{2(1 - \bar{R})} \tag{9}$$

Circular Standard Error: Circular standard error is a simple method which is used to determine the confidence interval when the sample size is ≥ 25.

The average of the data in the real part of the second trigonometric moment is calculated by the following equation

$$\bar{\alpha}_2 = \frac{1}{n}\sum_{i=1}^{n} cos2(\theta_i - \bar{\theta}) \tag{10}$$

and the standard error of mean direction is calculated (11) by the following equation

$$\sigma = \sqrt{\frac{n(1-\bar{\alpha}_2)}{2R^2}} \tag{11}$$

Concentration Parameter: The concentration parameter indicated by κ reflects whether the data set is homogeneously distributed on the circle or shows a concentration in the reference direction. It can be said that the data set is distributed uniformly on the circle when this value is 0 and that serious deviations from homogeneity occurred, in other words, data showed a concentration in the reference direction when it is greater than 2 (9). A suitable approach for the concentration parameter has been defined as the following

$$\hat{\kappa} = \begin{cases} 2\bar{R} + \bar{R}^3 + 5\bar{R}^5/6 & \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + \frac{0.43}{(1-\bar{R})} & 0.53 \leq \bar{R} < 0.85 \\ 1/(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}) & \bar{R} \geq 0.85 \end{cases} \tag{12}$$

Circular Distributions: They are probability distributions in the 0-2π range on the unit circle. Although there are many distributions on the circle, studies on very few of them were carried out. The most important ones of these distributions are the uniform distribution which is the most basic distribution on the circle, and the von Mises distribution which is expressed as the circular normal distribution, which plays an important role in statistical inference (3).

Uniform Distribution: If the observations are uniformly distributed on the circle, this distribution is expressed as a circular uniform distribution. In circular uniform distribution, all directions have equal probability between 0^o and 360^o. Since all of the observations are equally distributed on the circle, the mean resultant length is equal to 0 and the variance of the distribution is equal to 1. Therefore, the mean direction is undefined in the circular uniform distribution. The most important feature of this distribution is that it is not affected by rotation and reflection (2).

In uniform distribution; probability density function:

$$f(\theta) = \frac{1}{2\pi}, \qquad 0 \le \theta < 2\pi \tag{13}$$

Mean direction	μ: undefined
Mean resultant length	$\bar{\rho}$: 0
Circular dispersion	δ: ∞
p. cosine moment	α_p: $0, p \ge 1$
p. sine moment	β_p: $0, p \ge 1$ (9).

Circular Normal Distribution (von Mises): If circular random variable θ has a normal distribution, the distribution is expressed as the von Mises distribution. The most important distribution on the circle in terms of statistical inference is the von Mises distribution (12).

Bessel function converted from the first type zero order is

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} e^{\kappa \cdot cos(\theta-\mu)}\, d\theta \tag{14}$$

the probability density function of the von Mises distribution is

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cdot cos(\theta-\mu)} \tag{15}$$

and here, it is defined as $0 \le \theta,\ \mu \le 2\pi,\ \kappa \ge 0$.

The concentration parameter κ is a parameter that measures the concentration around the mean direction μ. Therefore, as the value κ increases, a high concentration also occurs around μ, which is the mean direction of the population (13). Furthermore, when von Mises distribution is $\kappa = 0$, it is converged to uniform distribution, for the small values of κ, it is converged to cardioid distribution, and when it is $\kappa > 2$, spiral is converged to normal distribution. The effect of $1/\kappa$ in circular normal distribution and the effect of σ^2 in normal distribution are almost the same.

In Von Mises distribution;

Mean direction	μ: $\bar{\theta}$
Mean resultant length	$\bar{\rho}$: $A_1(\kappa)$
Circular dispersion	δ: $[\kappa A_1(\kappa)]^{-1}$
p. cosine moment	α_p: $A_p(\kappa)$
p. sine moment	β_p:0, $p \ge 1$'dir (9).

Hypothesis Tests for Mean Direction: An important question of circular statistics is how observations are distributed on the circle. It is stated that the distribution is not homogeneous if most of the data do not show a concentration around the mean direction, and that the distribution is homogeneous if the data has a uniform distribution on the perimeter of circle (14).

One-Sample Mean Direction Test: The confidence interval at significance level of $(1-\alpha)$ can be used for one-sample mean direction test. It is necessary that the sample size should be at least 25 and the circular standard error should be determined for the hypothesis testing of mean direction and confidence interval. Whether the calculated mean direction is different from any mean direction given is tested.

The confidence interval for test statistics is calculated as the following

$$\mu_{gs} = \bar{\theta} \pm sin^{-1}\left(Z_{\frac{\alpha}{2}} \cdot \sigma\right) \quad (16)$$

If the mean direction is in the confidence interval given, hypothesis H_0 is accepted at a significance level of α (2).

Two-Sample Mean Direction Test: In this test, whether the mean directions of two circular distributions are different from each other is tested. The test statistics recommended by Watson and Williams is calculated (15) by the following equation

$$F_h = \left(1 + \frac{3}{8\kappa}\right)\left[\frac{(N-2)(R_1+R_2-R)}{N-(R_1+R_2)}\right] \quad (17)$$

Results

In this study, the representation methods of circular data, the calculation of descriptive statistics, mean direction, concentration parameters, hypothesis tests for compliance with uniform distribution and von Mises distribution, and hypothesis tests for the equality of mean directions, concentration parameters and distributions were introduced. Then, an application of these methods was performed with a data set including the times of birth of infants obtained from Van Yüzüncü Yıl University Faculty of Medicine Obstetrics and Gynecology Department.

The primary aim of the study was to introduce the circular data analysis which is not commonly used in practice and to show its applicability on a real data set. For this reason, in the discussion section of the study, biological interpretations of the results are not mentioned, and statistical interpretations are mainly emphasized.

Table 1. Circular descriptive statistics of the times of birth and hypothesis tests

n=179	Time (Hour) [360^o]	
Actual Mean Direction ($\bar{\theta}$)		321.938
Mean Resultant Length ($\bar{R}$)		0.0381
Circular Variance (V)		0.9619
Circular standard Deviation (v)		146.4416
Circular Dispersion (δ)		305.6192
Von Mises Concentration Parameter (κ)		0.0764
Skewness (s)		0.0561
Kurtosis (k)		-0.1048
Mean Cos ($\bar{C}$)		0.0300
Mean Sin($\bar{S}$)		-0.0235
Mean Cos ($\bar{C}_2$)		0.0281
Mean Sin($\bar{S}_2$)		0.1068
Mean Direction H_0=θ		
Score Test	Z= 0.1982	p= 0.6562
Likelihood Ratio Test	Z= 0.1981	p= 0.6563
Watson & Williams Test	F= 1.5013	p= 0.2221
Stephens Test	Z= 0.1982	p= 0.6562
Uniform Distribution Goodness-of-Fit Test	U^2=0.0792	p= 0.3840
Von Mises Distribution Goodness-of-Fit Test	U^2=42.8782	p= 0.0050

In Table 1, for the first trigonometric moments of the data, it was found that the mean cosine component ($\bar{C}$) was 0.03, the mean sine component ($\bar{S}$) was -0.0235, the mean resultant length ($\bar{R}$) was 0.0381 and the mean direction($\bar{\theta}$) was 321.938. For the second trigonometric moments, it was found that the mean cosine component ($\bar{C}_2$) was 0.0281, and the mean sine component ($\bar{S}_2$) was 0.1068. These values are used in the calculation of circular descriptive statistics and tests. These statistics are needed to perform various tests on Von Mises distribution parameters.

The mean direction ($\bar{\theta}$), one of descriptive statistics of the times of birth of infants, is the expression of the average value of the data distributed on the circle in degrees, and this value was calculated as 321.938°. Accordingly, the mean time of birth is approximately 21:28. The mean resultant length($\bar{R}$) is the mean length of the resultant of all observations and is a measure determining the concentration. This value is between the range of (0,1). A value close to 1 indicates a high concentration, and a value close to 0 indicates that there is no concentration and that the data is uniformly distributed around the circle. Since it is $\bar{R} = 0.0381$, it is seen that the data is uniformly distributed on the circle, in other words, the times of birth are not concentrated at any time of the day. V is the circular variance, and this value is descriptive of the spread in the data set. Circular variance takes value between the range of 0 and 1. In the sample, the fact that circular variance gets close to maximum with the value $V = 0.9619$, indicates a high spread of observations on the circle. Circular standard deviation (v) refers to deviations from the mean direction. It was calculated to be $v = 146.4416^o$. Another measure of dispersion (δ) based on the first and second central trigonometric moments is the measure of circular scattering, and this value was calculated as 305.6192. The fact that the concentration parameter(κ), which shows the concentration of circular data on the circle, was calculated as 0.0764 indicates that the data was uniformly distributed on the circle.

With respect to skewness and kurtosis parameters, the fact that the circular skewness value is found to be close to zero indicates a symmetric single-mode data set around the mean direction. The fact that the kurtosis value is less than 0 indicates that the distribution is more kurtic, flat and that the data is heterogeneous.

The hypothesis under the Von Mises assumption is established as

H_0: The mean direction under the circular normal distribution is equal to 0°

H_1: The mean direction under the circular normal distribution is different from 0°.

Under this assumption, score, Likelihood ratio, Watson & Williams and Stephens test statistics are given. According to the four tests given, hypothesis H_0 is decided to be accepted, and it can be said that the mean direction of the times of birth is not different from the direction 0^o at a significance level of $\alpha = 0.05$.

The Watson test statistic was used to test whether the times of birth were uniformly distributed over 24 hours a day or complied with the normal distribution. For uniform distribution, hypothesis H_0 is accepted and it can be said that the times of birth are uniformly distributed on the circle at a significance level of $\alpha = 0.05$. For Von Mises distribution, ($p<0.05$) hypothesis H_0 is refused. Thus, it can be said that the sample is not compatible with the circular normal distribution (von Mises distribution) at a significance level of $\alpha = 0.05$.

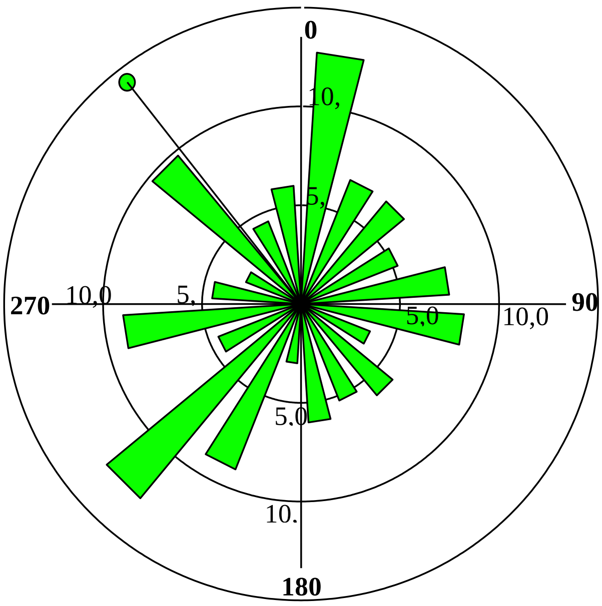

Figure 2: Rose diagram of the times of birth

The rose diagram for circular data on the times of birth is presented in Figure 2. When the figure is examined, it can be observed that the data is uniformly distributed on the circle, and that the mean direction is between 270° and 360°. When the range observed in the graphic is further reduced to a range of 305° to 335°, it can be said that the mean direction of the times of birth is approximately between 20:20 and 22:20.

Table 2. Circular descriptive statistics of the times of birth according to gender and hypothesis tests

n=179	Time (Hour) [360°]			
	Male (n=104)		Girl (n=75)	
Actual Mean Direction		285.0255		328.9989
Mean Resultant Length		0.0116		0.0788
Circular Variance		0.9884		0.9212
Circular standard Deviation		171.0192		129.1738
Circular Dispersion		3035.6413		67.451
Von Mises Concentration Parameter		0.0233		0.1588
Skewness		-0.1811		0.1742
Kurtosis		-0.0250		0.0638
Mean Cos		0.0030		0.0675
Mean Sin		-0.0112		-0.0406
Mean Cos		-0.0680		0.1614
Mean Sin		0.1663		0.0245
Mean Direction $H_0 = \theta$				
Equal Distributions		Test Statistics= 0.4842		p= 0.7850
Equal Directions		Test Statistics= 1.7478		p= 0.1879
Equal Concentration Parameters		Test Statistics= 0.3756		p= 0.5400
Score Test	Z= 0.0262	p= 0.8714	Z= 0.2476	p= 0.6188
Likelihood Ratio Test	Z= 0.0262	p= 0.8714	Z= 0.2474	p= 0.6189
Watson & Williams Test	F= 0.8973	p= 0.3457	F= 0.9036	p= 0.3449
Stephens Test	Z= 0.0262	p= 0.8714	Z= 0.2476	p= 0.6188
Uniform Distribution Goodness-of-Fit Test	U^2=0.067	p= 0.4636	U^2=0.0765	p= 0.4021
Von Mises Distribution Goodness-of-Fit Test	U^2=24.609	p= 0.0050	U^2=17.819	p= 0.0050

In Table 2, the data set of the times of birth was divided into two groups according to the gender of infants, and two samples were obtained by gender. The sample size was 104 baby boys and 75 for baby girls. The descriptive statistics and mean direction tests of these two samples, von Mises and normal distribution compliance tests, and the most common tests used to compare two groups were given to determine whether the times of birth varied by gender.

The value $\bar{\theta}$ calculated for the first sample indicates that the mean time of birth for baby boys is approximately 19:00. Since it is $\bar{R}$ =0.0116, it can be said that the data was uniformly distributed on the circle, in other words, the times of birth of baby boys were distributed over 24 hours of a day. The fact that the circular variance is close to maximum and the fact that the concentration parameter takes a small value like 0.0233 indicate that all observations were uniformly distributed on the circle. It was found that the circular standard deviation was 171.0192° and that the measure of circular dispersion, which is another measure of spread, was 3035.6413.

Although the skewness value of the first sample shows that the distribution is slightly skewed to the left, it can be said that this distribution is symmetric single-mode around the mean direction since it is close to zero. Similarly, although the kurtosis value also shows that the distribution is slightly flattened, it can be said that the circular kurtosis of distribution is the same as normal distribution since this value is close to zero.

The score, Likelihood ratio, Watson & Williams and Stephens test statistics were given to test whether the mean direction of the times of birth of baby boys was different from the H_0 mean direction (0°). According to these four tests given, hypothesis H_0 is decided to be accepted. Thus, it can be said that the mean direction of the times of birth of 104 baby boys (285.0255°) is not different from the direction 0^o at a significance level of $\alpha = 0.05$.

The Watson test was given to test whether the times of birth of baby boys were uniformly distributed or were compatible with the von Mises distribution. For uniform distribution, hypothesis H_0 is accepted. Accordingly, it can be said that the times of birth of baby boys were uniformly distributed around the circle at a significance level of $\alpha = 0.05$. For Von Mises distribution ($p<0.05$), hypothesis H_0 is refused. Therefore, it can be said that the sample is not compatible with the circular normal distribution at a significance level of $\alpha = 0.05$.

The value $\bar{\theta}$ calculated for the second sample indicates that the mean time of birth of baby girls is approximately 21:56. Since it is $\bar{R}$= 0.0788, it can be said that the data was uniformly distributed around the circle, in other words, the times of birth of baby girls were distributed over 24 hours of a day. The fact that the circular variance is 0.9212 and the fact that the concentration parameter was found to be 0.1588 indicate that the data was uniformly distributed around the circle. Furthermore, it was found that the circular standard deviation was 129.1738° and that the measure of circular dispersion was 67.451.

The skewness value given for girls indicates that the distribution is slightly skewed to the right. However, it can be said that the circular distribution is symmetric single-mode around the mean direction since this value is close to zero. It can be said that the kurtosis value is distributed slightly steeper than normal, but, the circular kurtosis of distribution is the same as normal distribution since this value is very close to zero.

The score, Likelihood ratio, Watson & Williams and Stephens test were given to test whether the mean direction of the times of birth of baby girls was different from the H_0 mean direction (0°). According to these four tests given, hypothesis H_0 is accepted. Thus, it can be said that the mean direction of the times of birth of baby girls (328.9989°) is not different from the direction 0° at a significance level of $\alpha = 0.05$.

The Watson test was used to test whether the times of birth of baby girls were uniformly distributed or were compatible with the von Mises distribution. For uniform distribution, hypothesis H_0 is accepted. Accordingly, it can be said that the times of birth of baby girls were uniformly distributed around the circle at a significance level of $\alpha = 0.05$. For Von Mises distribution ($p<0.05$), hypothesis H_0 is refused, and it can be said that the sample is not compatible with the circular normal distribution at a significance level of $\alpha = 0.05$.

The Score test was used for the equality of distributions, the Watson-Williams F-test was used for the equality of mean directions, and the Concentration homogeneity test was used for the equality of concentration parameters. According to the Score test, hypothesis H_0 was accepted, and thus, it can be said that both sample distributions are equal to each other at a significance level of $\alpha = 0.05$. According to the Watson-Williams F-test, hypothesis H_0 was accepted. Thus, it can be said that there is no difference between the mean directions of both samples at a significance level of $\alpha = 0.05$. According to the Concentration homogeneity test, hypothesis H_0 was accepted, and it can be said that sample concentration parameters are equal to each other at a significance level of $\alpha = 0.05$.

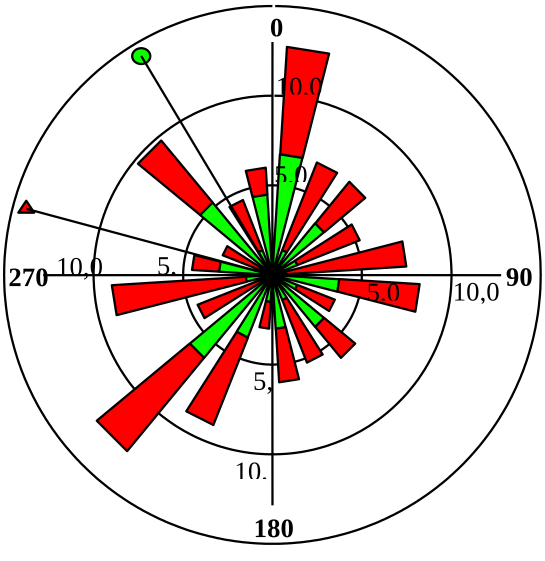

Figure 3: Rose diagram for the times of birth according to baby's gender

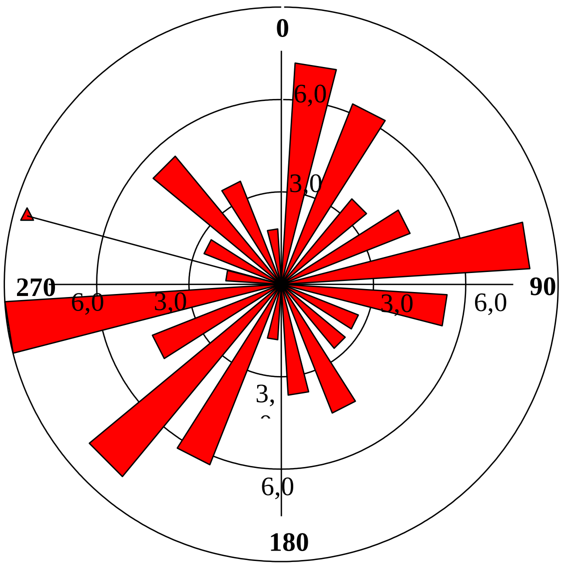

Figure 4: Rose diagram for the times of birth of baby boys

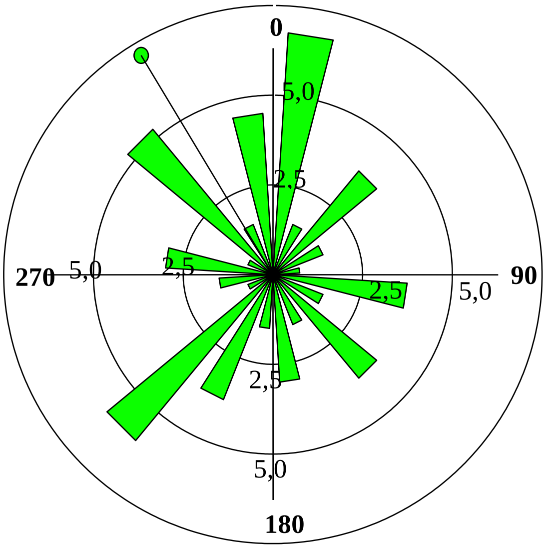

Figure 5: Rose diagram for the times of birth of baby girls

Baby boys are represented in red and baby girls are represented in green, and the rose diagram for the times of birth according to baby's gender is presented in Figure 3. The rose diagram for the times of birth of only baby boys is presented in Figure 4 while the rose diagram for the times of birth of baby girls is presented in Figure 5. When each graph given is examined separately, it can be said that the data was uniformly distributed on the perimeter of circle, in other words, there is no significant difference between the times of birth. The mean directions for the times of birth of baby boys and girls were observed between 270° and 360°. This range can be further narrowed and it can be said that there is no significant difference between the mean direction of the times of birth of baby boys and the mean direction of the times of birth of baby girls.

Discussion

Since the aim of the present study was to give information about circular data, descriptive statistics and hypothesis tests and to explain their functionality with an application, numerical values were not included in this section.

Since the validity of statistical data analysis methods to be used in a scientific research is directly related to the data structure and scale type, it is clear that the same methods cannot be applied to the data obtained from each scientific research, and therefore, it is necessary to select an analysis method which is appropriate to the scale type and the relevant data structure.

The fact that the topographic structures of the circle and the line are different from each other makes the data structures and scale types defined on them different from each other. Therefore, since the application of linear statistical methods to the data which is represented by an angle or is in the time cycle, in other words, on a circle or sphere surface will mostly produce false or misleading results, it is clear that it will be more accurate to use circular statistical methods in such periodic data.

It has been shown that there are significant differences between circular statistical methods and standard linear statistical methods, and that if a circular data set is attempted to be explained by standard linear statistical methods, the results may be inaccurate. Furthermore, the difficulties in the interpretation of descriptive statistics found by standard linear statistical methods were also mentioned. It is also necessary to emphasize that the results obtained will be meaningless in some cases.

It was pointed out that these important differences observed for descriptive statistics were also valid for test statistics and that the tests performed did not give proper results in some cases. Nevertheless, it was mentioned that the figures did not express a process since the point of time measured did not have an initial value no matter in which unit it is (such as seconds, minutes, hours, days, weeks, months, years).

Conclusion

it was shown that circular data cannot be analyzed by the analysis methods developed for standard linear data due to several reasons and that the results obtained in case of using standard methods could be inaccurate or could be meaningless in some cases, and therefore, it was emphasized that appropriate statistical methods should be used.

Conflict of Interest: The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Author's Contributions: **YD, ÖCB:** Research concept and design; data collecting, **YD:** Preparation of article, and Revisions. All authors approved the final version of the manuscript

Ethical issues: All Authors declare, originality and ethical approval of research. Responsibilities of research, responsibilities against local ethics commission are under the Authors responsibilities.

References

1. Lai MK.. Some Results on the Statistical Analysis of Directional Data. Master thesis, University of Hong Kong, Hong Kong Pokfulam, 1994.
2. Mardia KV, Jupp PE. Directional Statistics. John Willey and Sons. Inc., Chichester. 2000.
3. Jammalamadaka SR, SenGupta A. Topics in Circular Statistics. London: World Scientific; 2001. p.322.
4. Mahan RP. Circular Statistical Methods: Applications in Spatial and Temporal Performance Analysis. DTIC-ADA240751, Virginia, 53s. 1991.
5. Hintze JL. 2007. Circular Data Analysis. ss 927-955. Hintze, J.L. NCSS Help System, Published by NCSS, Utah 2823s. 2007.
6. Hussin AG, Jalaluddin JF, Mohamed İ. Analysis of Malaysian Wind Direction Data Using AXIS. Journal of Applied Science Research, 2006, 2(11): 1019-1021.
7. Ser G. Directional Data Analysis and An Application. Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi; Yıl: 2014 Cilt: 24 Sayı: 2; 121-126. 2014.
8. Berens P. CircSat: A MATLAB Toolbox for Circular Statistics. Journal of Statistical Software, 2009, 31: 1-21..
9. Fisher NI. Statistical Analysis of Circular Data. Cambridge University Press, New York, 277s. 1993.
10. Mardia KV. Statistics of Directional Data. Academic Press. Inc., London, 357s. 1972.
11. Brunsdon C, Corcoran J. Using circular statistics to analyse time patterns in crime incidence. Computers, Environment and Urban Systems, 2006, 30: 300-319..
12. Gaile GL, Burt JE. Directional Statistics. Concepts and Techniques in Modern Geography, London, 39s. 1980.
13. Peker KÖ, Bacanlı S. Descriptive Statistical Methods Applied to Circular Data and A Meteorological Application. Anadolu University Journal of Science and Technology, 2004, 5: 115-122.
14. Zar JH. Biostatistical Analysis. Prentice Hall, New Jersey, 944s. 2010.
15. Wheeler S, Watson GS. A distribution-free two-sample test on a circle. Biometrika, 1964,.51: 256-257.