Application of Circular ANOVA on Biological Data: Case Study on Crimean-Congo Hemorrhage Fever Cases in Turkey

Desta Firdu Mekonnen¹, Ensar Başpinar²

¹PhD candidate Ankara University Department of Biometry and Genetics
²Ankara University Instructor, Faculty of Agriculture

Corresponding Author: Desta Firdu Mekonnen
kenean001@hotmail.com

ABSTRACT

One of the priorities of Crimean-Congo hemorrhage fever management has been prevention methods. To apply these methods, understanding of transmission mechanisms and influencing factors is crucial. The main driving factor of Crimean-Congo hemorrhage fever time in a year. Plenty of analysis methods, like time series analysis, have been implemented to assess the relationship of Crimean-Congo hemorrhage fever transmission and time of the year. Nonetheless, the majority of analysis methods for the case lacked robustness.

Recently, the circular analysis in general and circular ANOVA, in particular, is gaining acceptance by many academicians to assess scenarios that have circularity in their nature. There are few stochastic circular models, but their applicability and tractability are not fully assessed as for their linear counterpart.

We examined circular ANOVA models on Crimean-Congo hemorrhage fever data presented by Dr. Fazilet DUYGU (2013) for Turkish public health organization titled “Dünya ve Türkiye’de Kırım-Kongo kanamalı ateşi epidemiyolojisi” to mean “Crimean-Congo hemorrhagic fever epidemiology in the world and Turkey.” Circular ANOVA method is used to assess the occurrence of the fever in different years. We transformed linear time in months to circular time and convert it to radian form since directional analysis is best suited to radian measures. We used Crimean-Congo hemorrhage fever cases in each month as a dependent linear variable and time in radian as a circular explanatory variable. We have calculated different descriptive statistics for the data and apply circular ANOVA on them and found very sensible methods with few drawbacks.

Keywords: Adjusted frequency, Circular Distribution, Circular ANOVA, Crimean-Congo hemorrhage fever.

1. INTRODUCTION

Crimean-Congo hemorrhage fever (CCHF) is one of the deadly hemorrhage fever that occurs in various regions of the globe. Particularly it is endemic to Africa, Asia, Eastern Europe and the Middle East (Appannanavar, and Mishra, 2011). It is a tick-borne zoonotic disease caused by CCHF virus. It has a widespread around the globe and creates a big challenge for nation economically weak. The distribution of the fever is facilitated by Hyalomma tick.

As one of tick-borne disease, CCHF incidence and transmission is influenced by geographical dynamics (Messina, J.P. et al., 2015).
From different geographical dynamics, climate and weather factors play big role for Hyalomma tick reproduction and hence CCHF incidence and transmission. It is a known fact geographical and climatic features of our planet is time dependent. There have been many researches to assess the relationship of CCHF and geographical dynamics in different parts of the world. Many models have been proposed for transmission and time dependency of CCHF.

Nevertheless, the majority of these analyses were based on linearity assumptions where they undermine the cyclical nature of the environmental and seasonal dynamics which depends on time. On the other hand, if we come to an agreement that CCHF is really influenced by time factors we have to comprehend the cyclical nature of time when the case occurs. That is where circular analysis arises.

There have been attempts to use circular analysis methods to apply for inferring the time of some scenarios occurs. However, the majority of these methods are time series like models that undermine the cyclical nature of time.

In this paper, we apply circular ANOVA models to assess the applicability of some of these models on Crimean-Congo hemorrhage fever data presented by Doğ Dr. Fazilet DUYGU (2016) for Turkish public health organization titled “Dünya ve Türkiye’de Kırım-Kongo kanamalı ateşi epidemiyolojisi” to mean “Crimean-Congo hemorrhagic fever epidemiology in the world and Turkey” Duygu (2016); in which data collection methods have been described in detail. Monthly malaria data were collected from the district for the 2008-2015 period.

For this study, we analyzed circular ANOVA to assess whether there are statistically significant differences between years of the CCHF distribution in Turkey in years 2008-2005 We used GetData Graphs digitizer 2.26 to extract data points from data plots and Excel software packages.

The method of obtaining data points from data plot using this software is quite simple. Load data plots using File menu, specify the X and Y minimum and maximum, click very carefully on to the data plots that we want data points, the result data points are shown on the right top corner of the page. Finally export data points into a format we want a file-export menu. The data points obtained in this way from data plots in Figure (1A) is presented in Table 1.

---

**Fig. 1** data plots and uploaded into GetData digitizer graphics.

4. DATA ADJUSTMENT

Our first step is to convert months into radian and adjust to the way one day represents 1 degree (0.017453 rad). Since months in a year do not have equal days, we have to adjust each month in the way to contribute equal frequencies. To do so, we have changed 365 days in a year to 360°.
First, we divide 30 by the number of days in each month we have the rate of each month’s contribution to 360°. 

**Adjustment** = \( \frac{30}{x} \); Where \( x \) is days of each month

We multiplied the adjustment result by the frequency of each month. To retain the total frequency, we added the adjusted frequencies and divided by the sum of frequencies before adjustment. Then, we multiplied the result by adjusted frequencies. The resulting frequencies are shown in Table 1.

5. RESULTS AND DISCUSSION

The goal of this study is to examine the relationship of CCHF disease in Turkey and occurrence time. The study variable is CCHF, and the independent variable is time in a circular sense.

The Research question is “is there any relationship between time of the years and occurrence of CCHF disease in Turkey.” For CCHF circular ANOVA analysis, Control and alternative hypotheses are developed as bellow

\[ H_0: \mu(\theta)_1 = \mu(\theta)_2 = \ldots = \mu(\theta)_k \]

\[ H_A: \text{Based on the occurrence of CCHF there is a difference at least between two circular times.} \]

As Rao (2001) pinpointed out under the assumption of same but unknown concentration parameter; Let \( (n_1, n_2, \ldots, n_k) \) denote the size of the combined sample directions. Under the null hypothesis, \( \mu \) denotes a common population mean direction and \( (R_1, R_2, \ldots, R\infty) \) are individual mean resultant vectors and \( R \) is an overall resultant vector based on \( n \) observations; \( \overline{R} \) can be computed as a median of individual \( R_i \)'s. The common mean resultant vector \( \overline{R} \) also can be calculated using vector addition rule. Calculation of common \( \overline{R} \) is presented in Appendix II.

Calculation of descriptive statistics of circular data is presented in many references (Rao, 2001; Mardia, and Jepp, 1972; Pewsey et al., 2013) calculation methods of circular descriptive statistics is presented in Appendix I at the end of this paper.

**Table 2 circular descriptive statistics of CCHF in Turkey**

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>Mean directions</th>
<th>MRV</th>
<th>Kappa</th>
<th>C var</th>
<th>Csd</th>
<th>rho2</th>
<th>rho1</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>1273</td>
<td>2.7329</td>
<td>0.8567</td>
<td>0.3325</td>
<td>0.6675</td>
<td>0.9780</td>
<td>-0.2227</td>
<td>0.0676</td>
<td>2.7187</td>
<td>2.7472</td>
</tr>
<tr>
<td>2009</td>
<td>1277</td>
<td>2.8633</td>
<td>0.8402</td>
<td>0.3265</td>
<td>0.6734</td>
<td>0.9859</td>
<td>0.0022</td>
<td>0.0532</td>
<td>2.8506</td>
<td>2.8759</td>
</tr>
<tr>
<td>2010</td>
<td>835</td>
<td>2.6477</td>
<td>0.8476</td>
<td>0.2488</td>
<td>0.7511</td>
<td>1.0992</td>
<td>-0.9481</td>
<td>0.0603</td>
<td>2.6311</td>
<td>2.6643</td>
</tr>
<tr>
<td>2011</td>
<td>1053</td>
<td>2.9268</td>
<td>0.8447</td>
<td>0.2903</td>
<td>0.7097</td>
<td>1.0364</td>
<td>0.0028</td>
<td>0.0420</td>
<td>2.9144</td>
<td>2.9392</td>
</tr>
<tr>
<td>2012</td>
<td>777</td>
<td>2.7217</td>
<td>0.8206</td>
<td>0.2287</td>
<td>0.7713</td>
<td>1.1320</td>
<td>0.0035</td>
<td>0.0261</td>
<td>2.7103</td>
<td>2.7330</td>
</tr>
<tr>
<td>2013</td>
<td>863</td>
<td>2.6789</td>
<td>0.8305</td>
<td>0.2493</td>
<td>0.7507</td>
<td>1.0985</td>
<td>0.0031</td>
<td>0.0309</td>
<td>2.6671</td>
<td>2.6906</td>
</tr>
<tr>
<td>2014</td>
<td>931</td>
<td>2.1658</td>
<td>0.8001</td>
<td>0.2530</td>
<td>0.7469</td>
<td>1.0926</td>
<td>0.0023</td>
<td>0.0319</td>
<td>2.1543</td>
<td>2.1772</td>
</tr>
<tr>
<td>2015</td>
<td>684</td>
<td>2.9139</td>
<td>0.8285</td>
<td>0.2099</td>
<td>0.7900</td>
<td>1.164</td>
<td>0.0043</td>
<td>0.0219</td>
<td>2.9028</td>
<td>2.9249</td>
</tr>
<tr>
<td>Common</td>
<td>7695</td>
<td>2.7206</td>
<td>0.5804</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Circular ANOVA**

Circular ANOVA is a test to compare more than two direction groups or two-time dependent measurement groups in a circular sense. Assumptions about the data, methods to follow and construction of ANOVA table is quite similar in logic with linear ANOVA. In linear ANOVA where group samples must be drawn from the normal distribution in circular case data must be drawn from Von Mises distribution, which is circular normal. In linear ANOVA where variables have to have a common variance which is a measure of

*To Cite This Article:* Desta Firdu Mekonnen, Ensar Baspinar. Application of Circular ANOVA on Biological Data: Case Study on Crimean-Congo Hemorrhage Fever Cases in Turkey. *International Annals of Medicine.* 2017;1(3). [https://doi.org/10.24087/IAM.2017.1.3.67]
homogeneity in circular case variables have to have common concentration parameter (kappa). In circular ANOVA majority of assumptions are still under review even some of them are not tested yet.

\( (n_i - \overline{R}_i) \) is one of measure of dispersions within a sample and \( (n - \overline{R}) \) is a measure of dispersion of a pooled sample.

Rao (2001) presented an identity between and within samples as follows:

\[
(n - R) = \left( n - \sum_{i=1}^{p} R_i \right) + \sum_{i=1}^{p} (R_i - R) 
\]

From the above identity also:

\[
(n - R) = \left( \sum_{i=1}^{p} (n_i - R_i) \right) + \sum_{i=1}^{p} (R_i - R)
\]

This is the breakdown of dispersion measures into:

Total dispersion = within sample dispersion + between samples dispersion

The approximate circular ANOVA table is presented as follows;

\[
df_B = p - 1 \quad SS_B = \sum_{i=1}^{p} (R_i - R) 
\]

\[
MS_B = \frac{\sum_{i=1}^{p} (R_i - R)}{p - 1}
\]

\[
df_w = \sum n_i - p \quad SS_w = \sum_{i=1}^{p} (n_i - R_i)
\]

\[
MS_w = \frac{\sum_{i=1}^{p} (n_i - R_i)}{\sum n_i - p}
\]

\[
df_T = \sum n_i - 1 \quad SS_T = n - R
\]

Then the F-value is calculated as:

\[
F = \frac{\sum_{i=1}^{p} (R_i - R) / (p - 1)}{\sum_{i=1}^{p} (n_i - R_i) / \sum n_i - p}
\]

Using calculation methods shown in Appendix I, individual mean resultant vectors of years 2008-2015 is presented in Table 4. Even if a number of observations in each group (years) are unequal, it has little influence on the calculation of common mean resultant vector, unlike linear ANOVA case where unbalanced observation size can affect the finding of common variance.

Where \( R = 0.580445 \) is a common mean resultant vector of a pooled data.

**Table 4 sources of variation summary**

<table>
<thead>
<tr>
<th>N_i</th>
<th>( \overline{R}_i )</th>
<th>( \overline{R}_i \cdot \overline{R} )</th>
<th>N_i ( \overline{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1273</td>
<td>0.856744</td>
<td>0.276299</td>
</tr>
<tr>
<td>2</td>
<td>1277</td>
<td>0.840117</td>
<td>0.259671</td>
</tr>
<tr>
<td>3</td>
<td>835</td>
<td>0.847609</td>
<td>0.267164</td>
</tr>
<tr>
<td>4</td>
<td>1053</td>
<td>0.844736</td>
<td>0.264291</td>
</tr>
<tr>
<td>5</td>
<td>777</td>
<td>0.820642</td>
<td>0.240197</td>
</tr>
<tr>
<td>6</td>
<td>863</td>
<td>0.830543</td>
<td>0.250097</td>
</tr>
<tr>
<td>7</td>
<td>931</td>
<td>0.800066</td>
<td>0.219621</td>
</tr>
<tr>
<td>8</td>
<td>684</td>
<td>0.82847</td>
<td>0.248025</td>
</tr>
<tr>
<td>SUM</td>
<td>7695</td>
<td>2.025364</td>
<td>7688.60</td>
</tr>
</tbody>
</table>

**Table 5 circular ANOVA calculation table**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degree of Freedom</th>
<th>Sum of Square</th>
<th>Mean of Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Samples</td>
<td>p-1</td>
<td>SSB</td>
<td>MSB</td>
</tr>
<tr>
<td>Within Samples</td>
<td>( \Sigma n - p )</td>
<td>SSW</td>
<td>MSW</td>
</tr>
<tr>
<td>Total</td>
<td>( \Sigma n - 1 )</td>
<td>SST</td>
<td></td>
</tr>
</tbody>
</table>

To reject the null hypothesis which claims indifference, p-value less than 0.05 is needed or the calculated F-value has to be greater than the critical F-table value for a specified degree of freedoms and I-type error value. Based on the calculated F-value (0.289288) < Ftable (7964, 7, 0.05) equals 3.23032.

Interpretation of Circular ANOVA Result

One way circular ANOVA was conducted to examine the relationship between years and occurrence of CCHF disease in Turkey. The result of ANOVA does not give enough evidence to reject the null hypothesis claims there is no difference between years in the occurrence of CCHF disease.

6. CONCLUSION

The main purpose of this study was to apply circular ANOVA on CCHF disease occurrence in Turkey. The circular analysis is very different in many aspects from a customary analysis methods we know. The basic difference is data management. In linear analyses where data magnitudes are analyzed in circular analysis time where observations occur is analyzed. In this paper we applied a circular ANOVA on occurrence time CCHF in Turkey from the year 2008-2015; data collected and presented by Doç. Dr. Fazilet DUYGU (2013) for Turkish public health organization. First, we extracted data points from data plot and found the time where these data points occur in radian.

The basic entity in circular ANOVA is not Mean direction, but it is Mean resultant vector. Due to this fact, we depended our analysis on calculating of mean resultant vectors of each year in detail and used it to construct ANOVA table.

The main problem of Circular ANOVA is lack specified posthoc analysis. After rejecting the null hypothesis claiming no difference, there are no comparison methods between groups. This makes Circular ANOVA is a test method urgently in need of “circular posthoc” methods otherwise the method loses its credibility.

Acknowledgement

We would like to acknowledge Doç. Dr. Fazilet DUYGU for being a basis for our study.

REFERENCES

Appendix I.

Mean Direction

We treated malaria case as grouped data and cases considered to happen in the middle of each month. We used the formula below to calculate grouped circular data.

\[
\bar{C}_n = \frac{1}{n} \sum_{i=1}^{k} f_i \cos(\theta_i) \quad \bar{S}_n = \frac{1}{n} \sum_{i=1}^{k} f_i \sin(\theta_i) \quad R = \sqrt{\bar{C}^2 + \bar{S}^2}
\]

where \(n = \sum_{i=1}^{k} f_i\)

\[
\cos(\bar{\theta}) = \frac{\bar{C}_n}{R} \quad \sin(\bar{\theta}) = \frac{\bar{S}_n}{R} \quad \bar{\theta} = \text{arctan}\left(\frac{\sin(\bar{\theta})}{\cos(\bar{\theta})}\right)
\]

The circular mean direction is quadrant dependent, and the actual mean direction should be specified based on what quadrant sine and cosine components fall.

\[
\begin{cases} 
\text{if } C_n > 0 \text{ and } S_n \geq 0; \mu_{dir} = \mu_{cat} \\
\text{if } C_n = 0 \text{ and } S_n > 0; \mu_{dir} = \frac{\pi}{2} \\
\text{if } C_n < 0; \mu_{dir} = \mu_{cat} + \pi \\
\text{if } C_n \geq 0 \text{ and } S_n < 0; \mu_{dir} = \mu_{cat} + 2\pi \\
\text{if } C_n = 0 \text{ and } S_n = 0; \mu_{dir} = \text{undefined}
\end{cases}
\]

Circular variance and standard deviation

\[
\text{Var} = 2(1 - \bar{R})
\]

In the circular variance above equation is to be able to get the small variation between circular variables otherwise, the variation would be unseen especially for data concentrated around the mean direction. Fisher (1995) presented circular standard deviation to be \(\sigma = (2\text{var})^{1/2}\). For the data concentrated around the mean direction. Circular standard deviation is found to be \(\sigma = (2\text{var})^{1/2}\). Fisher (1995) also used \(\delta = \frac{1 - \hat{\rho}_2}{2\sigma^2}\) as circular standard deviation where, \(\hat{\rho}_2 = \frac{1}{n} \sum_{i=1}^{n} \cos(\theta_i - \bar{\theta})\) to calculate confidence intervals.

Confidence interval

\[
\hat{\mu}(\theta) = \bar{\theta} \pm \sin^{-1}(Z_{\alpha/2}, \hat{\sigma}) \quad \hat{\sigma} = \frac{1 - \hat{\rho}_2}{2r^2} \quad \hat{\rho}_2 = \frac{1}{n} \sum_{i=1}^{n} \cos 2(\theta_i - \bar{\theta})
\]
Appendix II.

To calculate the common mean resultant vector, use individual mean resultant vectors and X (cosine) and Y (sine) components of mean direction. For CCHF data

Table Appendix II. Mean resultant vectors, cosine and sine components of mean directions

<table>
<thead>
<tr>
<th>R</th>
<th>Angle</th>
<th>X comp</th>
<th>Y comp</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.856744</td>
<td>2.732949</td>
<td>0.85577</td>
<td>0.130309</td>
</tr>
<tr>
<td>0.840117</td>
<td>2.863281</td>
<td>0.839068</td>
<td>0.143029</td>
</tr>
<tr>
<td>0.847609</td>
<td>2.64773</td>
<td>0.846704</td>
<td>0.122312</td>
</tr>
<tr>
<td>0.844736</td>
<td>2.926842</td>
<td>0.843634</td>
<td>0.149447</td>
</tr>
<tr>
<td>0.820642</td>
<td>2.721681</td>
<td>0.819716</td>
<td>0.129238</td>
</tr>
<tr>
<td>0.830543</td>
<td>2.678904</td>
<td>0.829635</td>
<td>0.125208</td>
</tr>
<tr>
<td>0.800066</td>
<td>2.16578</td>
<td>0.799494</td>
<td>0.081847</td>
</tr>
<tr>
<td>0.82847</td>
<td>2.913882</td>
<td>0.827399</td>
<td>0.148127</td>
</tr>
<tr>
<td>6.661421</td>
<td>1.029518</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\left(\bar{X}_{\text{comp}}\right)_i = R_i \cdot \cos(\bar{\theta})_i \\
\left(\bar{Y}_{\text{comp}}\right)_i = R_i \cdot \sin(\bar{\theta})_i \\
R_{\text{common}} = \sqrt{\sum (X_{\text{comp}})_i^2 + \sum (Y_{\text{comp}})_i^2}
\]