

Tutorial 2 - Circular descriptives

Getting started in R

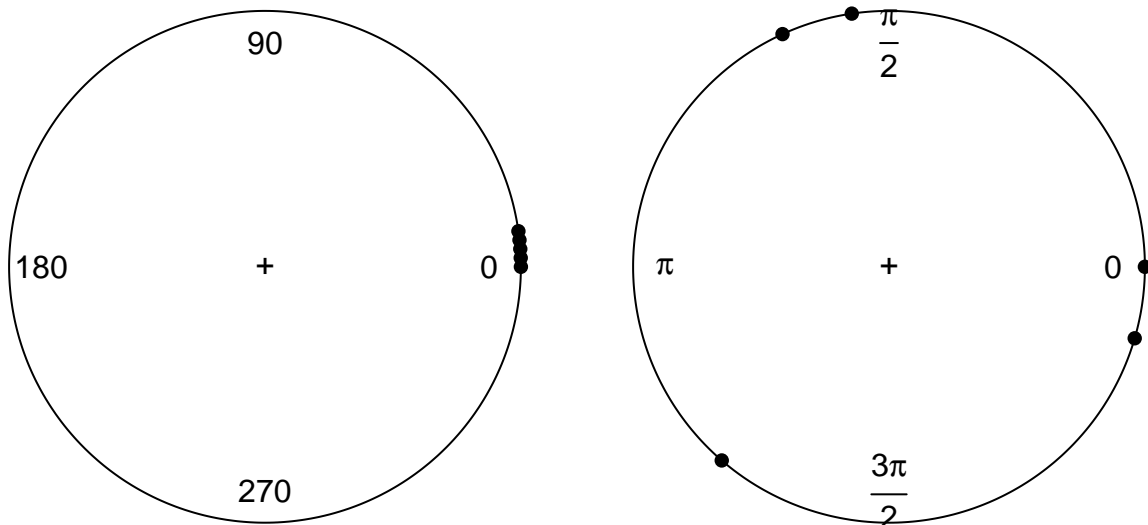
In the previous tutorial we discussed what circular data is and how it is measured. In this tutorial we will start analyzing circular data. For the analyses of circular data we will be using the open source programming language R. R is available for Windows, MacOS and Linux, and can be downloaded from <https://cran.r-project.org/>.

One tool that can be used to analyze and plot circular data is the package `circular` in R. To install and load this package the following code can be used

```
install.packages("circular")  
library(circular)
```

For circular data there are two units of measurement that are commonly used, degrees and radians. A circle has a total of 360 degrees, or equivalently, a total of 2π (≈ 6.28) radians. Values cannot exceed this total, and if they do most functions will automatically scale them back to the circular range (e.g. 361 degrees equals 1 degree). To showcase this, we specify one datasets containing the observations $[0, 2, 4, 6, 8]$. We can enter this data into R using the function `as.circular`. The argument `units` can be used to specify the measurement type in either degrees or radians.

```
data.deg <- as.circular(c(0, 2, 4, 6, 8), units = "degrees")  
data.rad <- as.circular(c(0, 2, 4, 6, 8), units = "radians")
```



The dataset in degrees (left) and the dataset in radians (right) plotted on the circle.

If the data is measured in degrees the observations will be concentrated on one side of the circle, as the circular range extends from zero to 360. If the data is measured in radians however, the data will be spread across the entire circular range of 2π . The value of 8 will exceed this maximum and is thus calculated as 1.72.

This shows that we must be very careful: the data entered in `as.circular` must match the units of observation. In R we can always convert our values from radians to degrees and vice versa using the function `conversion.circular`.

```
data.rad <- conversion.circular(data.deg, units = "radians")
```

The mean direction

A mean value in circular data is referred to as the *mean direction*. For example, the mean direction in a dataset with observations on a compass pointing north (0°) and east (90°) would be northeast (45°). Although intuitively this makes sense, it often is not as straightforward to calculate as the mean in linear data. Unlike in linear data, it is not possible to take the sum of all values and divide them by the total number of observations. If one would have a dataset with an observation of 350° and 10° then the mean direction on the circle will be drastically different from the linear mean.

Given a dataset with a total number of n observations θ , the mean direction of θ values requires us to calculate the sum of cosines ($C = \sum \cos\theta_i$), the sum of sines ($S = \sum \sin\theta_i$) and $R = \sqrt{C^2 + S^2}$, which is called the resultant length. The mean direction is then given by:

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

The mean direction can easily be computed in R. The package `circular` contains the function `mean.circular`, which integrates with the base R function for calculating the mean.

```
mean(data.rad)

## Circular Data:
## Type = angles
## Units = radians
## Template = none
## Modulo = asis
## Zero = 0
## Rotation = counter
## [1] 0.8584073
```

As we can see, the mean direction of values in radians are very different from the linear mean of $[0, 2, 4, 6, 8]$. Where the regular mean simply takes the sum of all values and divides them by the total amount of values, the mean direction takes account of the fact that the data is defined on a circular continuum.

Measures of concentration and spread

The mean direction is a measure of location, just like the mean in linear data. Similarly, we can also obtain descriptives that are a measure of spread or concentration. These indicate how much variation there is in observations, similar to the variance in linear data.

The mean resultant length (\bar{R}) is one such measure, and is obtained by $\bar{R} = R/n$, where $R = \sqrt{C^2 + S^2}$ and n is the number of observations. This is a measure of *concentration* that can take on values between zero and

one. The value 1 indicates that all values of the dataset are concentrated at a single location. A value close to zero indicates a higher spread in observed values. The mean resultant length can be calculated using the function `rho.circular`.

```
rho.circular(data.rad)
```

```
## [1] 0.2279162
```

A mean resultant length of $\bar{R} \approx .228$ indicates a somewhat low concentration.

Other than the mean direction and mean resultant length, we can also obtain a circular variance and a circular standard deviation. Since the mean resultant length is a measure of concentration and the variance is a measure of spread, the variance in circular data (V_m) is simply 1 minus the mean resultant length. Its interpretation is opposite to that of the mean resultant length. The function `var.circular` will give you this value, but it is also easily obtained by $1 - \bar{R}$.

```
var(data.rad)
```

```
## [1] 0.7720838
```

The standard deviation (v) is *not* obtained using the square root of the variance, but rather using $(-2 \log(1 - var))^{1/2}$. The function `sd.circular` in R provides this value.

```
sd(data.rad)
```

```
## [1] 1.719754
```

Please note the circular package always returns this value in radians. We can obtain a standard deviation in degrees if desired using `conversion.circular`.

```
conversion.circular(sd(data.rad), units = "degrees")
```

```
## Circular Data:  
## Type = angles  
## Units = degrees  
## Template = none  
## Modulo = asis  
## Zero = 0  
## Rotation = counter  
## [1] 98.53466
```

The standard deviation in our dataset is approximately 1.72 radians or 98.5 degrees.

When working with circular data one peculiar situation can arise in which both the mean direction and the mean resultant length are unreliable or even undefined. When data is distributed on the circle in such a manner that there are two opposite modes, the mean direction will be largely meaningless. Furthermore, the mean resultant length will be close to zero, indicating complete dispersion of datapoints where in practice this is not the case. On rare occasions it is possible that all values are cancelled out by opposite values, resulting in an undefined mean direction. For example, what is the mean direction of north and south: east or west? In such a situation the mean resultant length will equal zero. It is therefore always recommended to plot the distribution of circular data before drawing any conclusions about the concentration or location of datapoints.

Conclusion

The mean direction ($\bar{\theta}$), mean resultant length (\bar{R}), variance (V_m) and standard deviation (v) are ways to describe location and spread in circular data. These measures are however not always reliable. For that reason, it is recommended one always plots their data and inspects it visually as well. In the next tutorial, we will discuss why these measures are not always reliable by looking into circular distributions. We will also cover how to plot circular data in R.

References

The equations used in this tutorial can, for example, be found in the book by Fisher (chapter 2.3). Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge University Press.