

Circular Statistics: A short description of SAS-macros written by:

Mathias Kölliker

Zoological Institute, University of Basel, Evolutionary Biology
Vesalgasse 1, 4051 Basel, Switzerland

When your variables of interest are distributed on a circle (e.g., angles, times of day, dates) linear statistics often cannot be used properly. The major problem with circular variables is that values which are far apart on a linear scale can be very close together in reality. Take for example an angle of 10 degree. On a linear scale, it is 340 degrees from a second angle of 350 degrees. The true distance between the two angles, however, is only 20 degrees. This causes problems when using traditional statistical tests. Commercial software packages often do not provide tools for circular statistics, making circular statistical analyses often hard and lengthy to perform.

Because I was interested in orientation issues of great tit (*Parus major*) nestlings (Kölliker & Richner, 2004) I became interested in circular statistics and wrote a few SAS-macros covering some basic needs when doing statistics with circular variables. The macros are based on the highly recommendable textbooks on circular statistics by Batschelet (1981) and Fisher (1993). Please consult these texts for understanding the background of circular statistics.

Circular statistics macros:

You find the collection of the SAS-macros in a single Text-file (“circular_stats.sas”) downloadable on www.evolution.unibas.ch/koelliker/misc.htm). You need access to SAS to be able to use the macros. To run the macros, download the file and run it in SAS using the “%INCLUDE” statement:

```
%INCLUDE "path\circular_stats.sas";
```

“path” is the physical location of the “circular_stats.sas” file on your machine (e.g., c:\Documents and Settings\My Computer\My Documents\....).

See details below on how to launch the macros. You can also download from the website “www.evolution.unibas.ch/koelliker/misc.htm” an example data-set and template code that you may use, test on your computer and adjust to your own needs.

Importantly, the macros assume that you imported your data-set to SAS prior to running the macros (import is included in the template code that can be downloaded). Further, it is easiest to define the macro-variables if you work with SAS libraries (i.e., using the LIBNAME statement).

The SAS macros cover the following analyses and tests:

1: Circular moments

The macro calculates sample size, mean vector length, mean angle, angular variance, angular deviation, circular standard deviation for a circular variable.

How to start the macro: Run the following line, adjusted to your data definitions, in the SAS-editor after running the above “%INCLUDE” statement:

```
%circular_moments(data_in, data_out, circvar, arcs, rad_deg, remarks);
```

data_in: provide here the full path-information of your data file. The macro uses this information to access the to-be-analyzed data.

data_out: here you can specify a path for an output SAS-data set containing the results.

circvar: the name of your variable of interest.

arcs: Sometimes, angles are measured using sectors – arcs – on a circle (e.g., 12 arcs of 30 degrees each). Provide here the number of arcs you are using. This is important because in such cases a correction is required. If your circular-variable is based on accurate measures, enter “no” here.

rad_deg: specify here whether your variable is measured in radians (rad) or degrees (deg).

remarks: add whatever information you think is important to understand the output, e.g., a spelt out name of your variable.

2: Raileigh-Test for uniform distribution

A randomization procedure (permutation test) is used to approximate a p-value. A significant p-value indicates significant deviation from a uniform distribution, hence concentration of the data in a certain direction. The test assumes a uni-modal distribution and cannot be used to test for multi-modal deviation from a uniform distribution.

How to start the macro: Run the following line, adjusted to your data definitions, in the SAS-editor after running the above “%INCLUDE” statement:

```
%rayleigh_test(data_in, data_out, circvar, arcs, rad_deg, remarks,  
iterations = 5000);
```

Definitions of the macro-variables are as above with one exception:

iterations = : specify here the number of permutations you want to use to approximate a p-value. More is always better, but it can take time. It is sometimes recommended to use at least 5000 permutations, which is usually feasible in suitable time. Default is no iterations.

3: Circular correlation

Four different correlation coefficients are calculated following propositions by different authors. Three of them (following are parametric (analogous to the Pearson product-moment correlation coefficient in linear statistics), and one is non-parametric (analogous to the Spearman correlation coefficient in linear statistics). P-values are approximated by use of a randomization procedure (permutation test).

How to start the macro: Run the following line, adjusted to your data definitions, in the SAS-editor after running the above “%INCLUDE” statement:

```
%circular_corr(data_in, data_out, circvar1, circvar2, rad_deg,  
iterations = 5000, output = yes);
```

Definitions of the macro-variables are as above except for the requirement of two circular variables (the two you want to correlate). The permutation test is slow here because four correlation coefficients are calculated simultaneously at each step. Thus, kick of the macro and go have a cup of coffee.....!

4: Two-sample test (Watson-Williams test)

This macro allows you to test if the mean directions of two samples differ significantly. The p-value is approximated using a randomization procedure (permutation test).

How to start the macro: Run the following line, adjusted to your data definitions, in the SAS-editor after running the above “%INCLUDE” statement:

```
%two_sample_test(data_in, data_out, circvar, groupvar, rad_deg,  
iterations = 5000);
```

Definitions of the macro-variables are as above except for the grouping variable (groupvar) corresponding to your factor. Note that the two groups you want to compare need to be specified in your data set by the numerals 1 and 2, respectively.

Note on p-values:

It is possible that you are getting a p-value of “.000” from the permutation test. This value cannot be interpreted as truly zero (p-values are never truly zero), but rather means that the permutation-test over the number of steps that you defined never generated a test-statistic as extreme as the one of your real data. Also, the number of digits is set to three for the p-value. You may report such a result as “p < 0.001”.

Reference:

If you use these macros for analyses that you report in a scientific publication, please cite Kölliker & Richner (2004) as the reference to these macros.

These macros are for free non-commercial and scientific use. If you have any question, comments, or if you find a bug, please contact me: mathias.koelliker-at-unibas.ch.

Have fun doing circular statistics!

Basel, August 2005

Mathias Kölliker

References

- Batschelet, E. 1981. *Circular statistics in biology*. London: Academic Press.
- Fisher, N. I. 1993. *Statistical analysis of circular data*. Cambridge: Cambridge University Press.
- Kölliker, M. & Richner, H. 2004. Navigation in a cup: chick positioning in great tit, *Parus major*, nests. *Animal Behaviour* 68: 941-948.