Hacettepe Journal of Mathematics and Statistics
Volume 43 (1) (2014), 97 – 115

POWER STUDY OF CIRCULAR ANOVA TEST AGAINST NONPARAMETRIC ALTERNATIVES

Feridun Tasdan * and Ozgur Yeniay †

Received 24 : 12 : 2012 : Accepted 17 : 10 : 2013

Abstract

This study compares circular ANOVA against bootstrap test, uniform scores test and Rao's test of homogeneity which are considered nonparametric alternatives. Circular ANOVA is one-way analysis of variance method to test the equality of mean directions in circular data analysis, but it requires some assumptions. The main assumption for circular ANOVA is that all r-independent samples must come from von Mises distribution with equal directional means and equal concentration parameters. On the other hand, nonparametric alternatives are distribution free methods and, therefore, does not require having von Mises distribution or equality of parameters. Literature of circular statistics is very limited on the comparison of these tests; therefore, a power simulation study is performed to compute the power of circular ANOVA against the nonparametric alternatives under assumptions of von Mises and non-von Mises populations. Power simulation study shows that bootstrap and uniform scores tests perform slightly better than circular ANOVA if the common concentration parameter, κ, is less than 1 under the assumption of von Mises distribution. If $\kappa \geq 2$, then bootstrap and circular ANOVA perform better than the other alternatives. Rao's test of homogeneity requires very large samples in order to reach the same power levels of competitive tests in this study. Finally, uniform scores tests performs better than circular ANOVA and bootstrap test if the sample sizes are small and the data comes from mixed von Mises distributions or wrapped Cauchy.

Keywords: Keywords:Bootstrap, Circular Data, Circular ANOVA, von Mises Distribution, Seasonal Wind Directions, Uniform Scores Test, Rao's Test.

2000 AMS Classification:

*Department of Mathematics Western Illinois University Macomb, IL 61455
Email: (F. Tasdan) F-Tasdan@wiu.edu

†Department of Statistics Hacettepe University Beytepe, Ankara, Turkey.
Email: (O. Yeniay) yeniay@hacettepe.edu.tr

1. Introduction

The history of circular data problems, which can be seen in biology, geography, medicine, meteorology, oceanography and many other fields, goes back to the 1950s, but we have seen more publications in the last 25 years. Several textbooks and many papers have been published in recent years about the circular data problems. [11], [16], [6], [9],[2] are excellent resources for circular data problems. Technological developments in computers and programming made it possible to analyze large or complicated circular data problems. There are several computer programs currently available for the analysis of circular data problems. One of these is R program with circular package, which is jointly developed by [1]. It is called "circular" in R package repository. In fact, some of the results in this study are obtained from this circular package.

Circular data is obtained by measuring directions or arrival times of subjects with respect to a reference point on the unit circle. This reference point or the choice of the origin is arbitrary and the final conclusions should not depend on it. For example, North can be taken as a reference point (considered as 0 degrees) on the unit circle. Therefore, circular data will have a domain of $[0, 2\pi)$ in radians or $[0, 360)$ in degrees depending on the definition of the problem. If the arrival times of patients to an emergency room are the main interest, then the data can be recorded in 24 hour clock notation (domain of [0:00, 24:00)) and can later be converted to the angles on the unit circle.

Moreover, two or more sample circular data problems have been increasingly common in recent years. Watson and Williams ([17]) introduced a test for the equality of r-population means in circular data problems. This test can be considered an equivalent of one-way ANOVA in the traditional linear data problems. In later years, [11] and [14] modified the Watson-Williams test for certain conditions, which are given in Section 2. Nonparametric tests are also developed for two or more sample circular data problems. The test of homogeneity of r-populations is proposed by [11] and [18]. It is called uniform scores test or Mardia-Watson-Wheeler test in the literature. The test is based on ranks of the combined samples, but it is very sensitive to the existence of ties. [2] suggested that Mardia-Watson-Wheeler test should not be used if there are many ties in the data, but a few ties could be broken by a randomization or average methods. [13] introduced a nonparametric test called "Rao's test of homogeneity" for the equality of r-populations (homogeneity of populations). The details of the test are given in Section 3. Also, a bootstrap based test for the equality of r-population means is available and promoted by [6] especially if the sample sizes are less than 10 or assumptions do not meet in circular ANOVA test. The next section will give some insight about the multi-sample method called circular ANOVA in circular statistics.

2. Circular ANOVA

Circular ANOVA (One-Way Analysis of Variance) has been proposed by [17] and later modified by [11] based on suggestions by [14]. The theory of circular ANOVA is discussed extensively by [6], [11] and [9] on pages 125-128. In an another important paper, [8] also discuss the drawbacks of suggestions by [14]. The first assumption of the circular ANOVA is that all random samples should come from von Mises distribution with a common concentration parameter κ such that $H_0 : \kappa_1 = \kappa_2 = .. = \kappa_r = \kappa$ (test of homogeneity of kappa). If the assumptions of having von Mises distribution and the test of homogeneity of the kappa parameters fail, then [6] proposes nonparametric approaches for the analysis of two or more samples in circular data. If the sample sizes $n_1, .., n_r$ are less than 25, the bootstrap approach is heavily emphasized by [6]. There are several options (analogous to Levene's test in linear data) available for testing that all κ parameters are equal. We will introduce one of them in the next section when we

Source	DF	SS	MS	F
Between Samples	$r-1$	$\sum R_i - R$	$(\sum R_i - R)/(r-1) = I$	$F_t = I/II$
Within Samples	$N-r$	$N - \sum R_i$	$(N - \sum R_i)/(N-r) = II$	
Total	$N-1$	$N-R$		

perform the large sample example with R's circular package. There is a necessity that either the common concentration parameter κ is given or must be estimated from the data. So, [6] proposes $\hat{\kappa} = median\{\hat{\kappa_1}, \hat{\kappa_2}, .., \hat{\kappa_r}\}$ as an estimator of κ if it is unknown. Depending on the value of the common concentration parameter, there are several alternative approaches for circular ANOVA. [6] categorizes these approaches in three sections: $\kappa \geq 2$, $1 < \kappa < 2$, and $\kappa \leq 1$.

First, assume that $\kappa \geq 2$ and state the hypothesis that

$$H_0 : \mu_1 = \mu_2 = .. = \mu_r \text{ vs } H_1 : \text{At least two are distinct.}$$

Let θ_{ij} (for $i = 1, .., r$ and $j = 1, .., n_i$) shows angular observations coming from a circular distribution on the unit circle. Let R be the resultant length of all $N(N = n_1 + n_2 + .. + n_r)$ observations. The variable R can be computed by using all observations $(\theta_1, ..., \theta_N)$ or [6] provided the following formula that uses individual sample resultant lengths $(R_1, R_2, ..R_r)$ and mean directions $(\overline{\theta_i})$. Let

$$R = [(\sum_{i=1}^{r} R_i \cos(\overline{\theta_i}))^2 + (\sum_{i=1}^{r} R_i \sin(\overline{\theta_i}))^2]^{0.5} \tag{2.1}$$

The test statistic for circular ANOVA is defined by

$$F_t = (N-r)(\sum_{i=1}^{r} R_i - R)/[(r-1)(N - \sum_{i=1}^{r} R_i)] \tag{2.2}$$

where F_t has an F distribution with r-1 and N-r degrees of freedoms. We reject the test if $F_t > F_{r-1,N-r}$. One advantage of this test is that the F critical values can be found in many statistics books. [11] defined a circular ANOVA table summarizes the result:

If $1 < \kappa < 2$, [14] proposes a modified test that uses correction a factor and it is defined as $F_t' = [1 + 3/(8 * \hat{\kappa})]F_t$. If $\kappa \leq 1$, then [11] proposes an approximate likelihood ratio test which is defined below,

$$-2\log_e \lambda \doteq \frac{2}{N}\{(\sum_{i=1}^{r} R_i)^2 - R^2\} = U. \tag{2.3}$$

where for a large N, U has an approximate chi-square (χ^2) distribution with $r-1$ degrees of freedom when H_0 is true. The expression for λ can be derived from the equation (2.3). Details of this approximation can be seen in [11] on page 164.

3. Nonparametric Tests

Recall that circular ANOVA is discussed in Section 2 and requires multiple assumptions: (i) r-samples are coming from (at least approximately) von Mises distribution, (ii) the concentration parameters (κ) are equal, (iii) the value of the common concentration parameter is larger than $\hat{\kappa} > 1$. In many real life situations, one or more of these assumptions may not be satisfied. Therefore, alternative tests for circular ANOVA must be considered in order to avoid those assumptions or replace circular ANOVA if the assumptions are not satisfied. Bootstrap test is one approach that avoids these assumptions listed above. Mardia-Watson-Wheeler test (also called uniform scores test) and Rao's test of homogeneity are also nonparametric tests that they do not require having von Mises distribution assumption or the equality of parameters. One disadvantage for Rao's test is that it requires sufficiently large sample sizes. These nonparametric alternatives are discussed in the following sections.

3.1. Bootstrap Test. The bootstrap method was first introduced by [3] and became popular in recent years due to technological advances in the computer sciences. With the bootstrap method, the original sample is treated as the population and a resampling procedure is performed on it. This is done by randomly drawing a sample of size n from the original sample (size n) with replacement. [4] introduced many bootstrap methods as an alternative to parametric methods. [5] and [7] studied bootstrap method for circular data problems extensively. An algorithm based on bootstrap test for circular data has also been discussed by [12]. They showed that the bootstrap based hypothesis testing method to test the equality of peak months for fish populations could be used by considering the months as circular variables. In comparison to the circular ANOVA, bootstrap test approach uses the bootstrap estimate of the test statistic (F statistic) from the combined samples of circular data. In each bootstrap step, bootstrap estimate of the test statistics ($F^\star$) is found and compared with the original test statistic which is computed from the original samples. Then an estimated significance value (p-value) of the bootstrap test is calculated by first finding the number of bootstrap test statistic which is greater than the original test statistics and dividing the result with the number of bootstrap runs (B replications). If the estimated significance value is less than or equal to level of significance, it means that there is a significant difference among the population mean directions and, therefore, $H_0 : \mu_1 = \mu_2 = ... = \mu_r$ is rejected.

The following bootstrap test algorithm can be defined in order to obtain the bootstrap significance value or p-value. The algorithm is somewhat similar to [6]'s definition of the bootstrap test for two or more samples but the main difference is that [6] does not combine the samples whereas the proposed bootstrap test combines the samples to create one large sample and draws a bootstrap sample from this combined sample, then partitions it into $n_1, n_2, ..n_r$ sub-samples randomly. Of course, bootstrap test is performed under H_0. Therefore, combining r-samples to create one large sample and re-sampling from this large sample is used in the proposed algorithm.

An algorithm for the construction of bootstrap test and finding p-value as follow:

(1) Let θ_{ij} for $i = 1, .., n_j$, and $j = 1, .., r$ be the angular measurements from $n_1, ..n_r$ samples. Calculate F_t test statistics using the original samples with "aov.circular" function in R.
(2) Draw a bootstrap sample of size $N = n_1 + n_2 + .. + n_r$ from the combined sample of θ_{ij} with replacement. Assign first n_1 observation to first level 1, then n_2 observations to level 2, and the last n_r observations to level r. This way $n_1, n_2, .., n_r$ observations are assigned to $1, 2, .., r$ samples respectively. Calculate the test statistics $F_b^\star$ using these samples.
(3) Repeat the last two steps for $b = 1, ..., B$.
(4) There are now $F_1^\star, .., F_B^\star$ estimated bootstrap test statistics.
(5) Find the number of $F_b^\star \geq F_t$ and then divide the result by B. The result gives $\hat{p} = \#\{F_b^\star \geq F_t\}/B$.
(6) Compare $\hat{p}$ by the level of significance α. If $\hat{p} \leq \alpha$, reject $H_0 : \mu_1 = \mu_2 = ... = \mu_r$. Otherwise, do not reject H_0.

3.2. Uniform Scores Test. A nonparametric test for the equality of two circular distributions is first presented by [18]. A few years later, two-sample case has been extended to k-sample case by [10]. For this reason, k-sample uniform scores test has also been called as Mardia-Watson-Wheeler test in the literature. The null and alternative hypothesis of the test is

H_0:All samples come from the same population
H_1:At least two are distinct.

Let θ_{ij} (for $i = 1, .., r$ and $j = 1, .., n_i$) show the combined samples of $n_1, n_2, .., n_r$, where each sample consists of angular observations on the circle. The testing procedure assigns ranks to all θ_{ij} and finds a uniform score or circular rank for each θ_{ij} as,

$$d_{ij} = \frac{2\pi(r_{ij})}{N} \text{ for } i = 1, .., r \text{ and } j = 1, .., n_i$$

where r_{ij} is the rank of jth observation from ith sample and $N = n_1+, .., +n_r$. A starting point should be set on the circle in order to find the ranks which can be assigned clock wise or counter clock wise on the circle. In fact, the test is invariant under all rotations as shown by [11], therefore the initial rank could be given to the smallest angle in the data. The test statistics is defined as

$$W = 2\sum_{i=1}^{r}(C_i^2 + S_i^2)/n_i \tag{3.1}$$

where

$$C_i = \sum_{j=1}^{n_i}\cos(d_{ij}) \quad and \quad S_i = \sum_{j=1}^{n_i}\sin(d_{ij}). \tag{3.2}$$

are the components of resultant vector for each sample. We should keep in mind that $\sum_{i=1}^{r} C_i = 0$ and $\sum_{i=1}^{r} S_i = 0$, where they could be used to check if the computations are correct in the formulas above. The test statistic, W, has an approximate chi-square (χ^2) distribution with degrees of freedom of 2(r-1) as shown by [10]. Therefore, if $W > \chi^2_{\alpha,2(r-1)}$, H_0 is rejected in favor of H_1. [6] suggests that this test is applicable if $n_i > 10$ for $i = 1, .., r$. Otherwise, a permutation test should be applied.

3.3. Rao's Test of Homogeneity. [13] proposed a test of homogeneity that it is considered large sample alternative of circular ANOVA test. The test is available from R circular package. The requirements to apply Rao's test of homogeneity tests is that the data must be unimodal and the sample size must be sufficiently large.

Let θ_{ij} (for $i = 1, .., r$ and $j = 1, .., n_i$) show the combined samples of $n_1, n_2, .., n_r$. Let X_i and Y_i denote the means of cosine and sine values for ith sample of size n_i such that

$$X_i = \frac{\sum_{j=1}^{n_i} cos\theta_{ij}}{n_i} \text{ and } Y_i = \frac{\sum_{j=1}^{n_i} sin\theta_{ij}}{n_i}$$

and $T_i = \frac{Y_i}{X_i}$ with asymptotic estimated variance of s_i^2 in which the details can be found in [13]. The test statistics, H, is defined as

$$H = \sum_{i=1}^{r} \frac{T_i^2}{s_i^2} - \left(\sum_{i=1}^{r} \frac{T_i^2}{s_i^2}\right)^2 / \left(\sum_{i=1}^{r} \frac{1}{s_i^2})\right)$$

Under H_0 and some general conditions, the test statistics H has a χ^2 distribution with $df = r-1$. For large values of H, the null hypothesis H_0 is rejected which implies different mean directions.

4. Large Sample Example

4.1. Application of Circular ANOVA. The city of Ankara is the capital of Turkey and has a population 4.4 million according to Turkish Institute of Statistics. The city has an elevation of 3077 feet (938 meters) and located at the central part of Turkey. Turkish State Meteorological Services (TSMS) has regional stations that collect and distribute weather related data in Turkey. The literature review did not reveal any studies about the analysis of the seasonal wind directions for the city of Ankara. This study will be the first in this regard. The data provided by TSMS consisted of daily wind directions of Ankara for the year of 2010. First, using the data provided, descriptive summary results were obtained for each seasons (winter, spring, summer and fall). Table 1 shows the descriptive statistics for four seasons. To see the seasonal differences, the data is divided

Table 1. Descriptive Statistics for Seasonal Wind Directions in Ankara

Parameters	Winter	Spring	Summer	Fall
Sample Size	90	92	92	91
Mean Direction(degrees)	108.38	140.93	111.48	116.79
Mean Resultant Length	0.6182	0.6458	0.7086	0.6727
Circ. Variance	0.3818	0.3542	0.2914	0.3273
Circ. Std. Deviation	0.9808	0.9464	0.8396	0.8963
Median Direction(degrees)	100.5	140.5	103	107

into four seasons(winter, spring, summer and fall), and rose diagrams(equivalent of histogram) are graphed for each season. Figure 1 shows the seasonal distribution of the wind directions for the year of 2010 in Ankara. In Figure 2, QQ plots of von Mises distribution for each season is shown. It is safe to assume that seasonal wind directions of Ankara (at least for the year of 2010) follow von Mises distribution.

Before performing a circular ANOVA test, we needed to find MLE of κ parameter for all four seasons. The common κ is estimated by $\hat{\kappa} = 1.693012$ with all the samples combined together. If we use [6]'s approach by finding the median of the four seasons, we find that $\hat{\kappa} = 1.754571$. Both results are very much comparable and on the interval $1 < \hat{\kappa} < 2$. See Table 2.

Assumption of the homogeneity of concentration parameters (κ) must be tested in the next step. The circular ANOVA test proposed by [17] assumes that all r concentration parameters are equal to the common concentration parameter κ. So,

$$H_0 : \kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = \kappa \text{ vs } H_1\text{: At least two are distinct.}$$

This must be tested before starting circular ANOVA method. The following results are obtained from R software using the package called "circular" and using "rao.test" function. The hypothesis test checks the equality of the concentration parameters, the results are from R software (See Table 3).

As we see from the result, the p-value of the test is 0.6171 which is greater than a level of significance of $\alpha = 0.05$ or even 0.10. Therefore, it is safe to assume that all concentration parameters are equal. Since the estimated common concentration, $\hat{\kappa}$, is between 1 and 2, we must use the modified F-test in circular ANOVA according to [6].

Table 2. κ parameter estimates for all four seasons. Table also includes common κ estimates which are the last two values

Winter	Spring	Summer	Fall	Common $\hat{\kappa}$	Fisher's $\hat{\kappa}$
1.585405	1.679172	2.024359	1.829970	1.6930	1.7545

Table 3. Test of Homogeneity of Kappa Parameter

df	ChiSq	P-value
3	1.79	0.6171

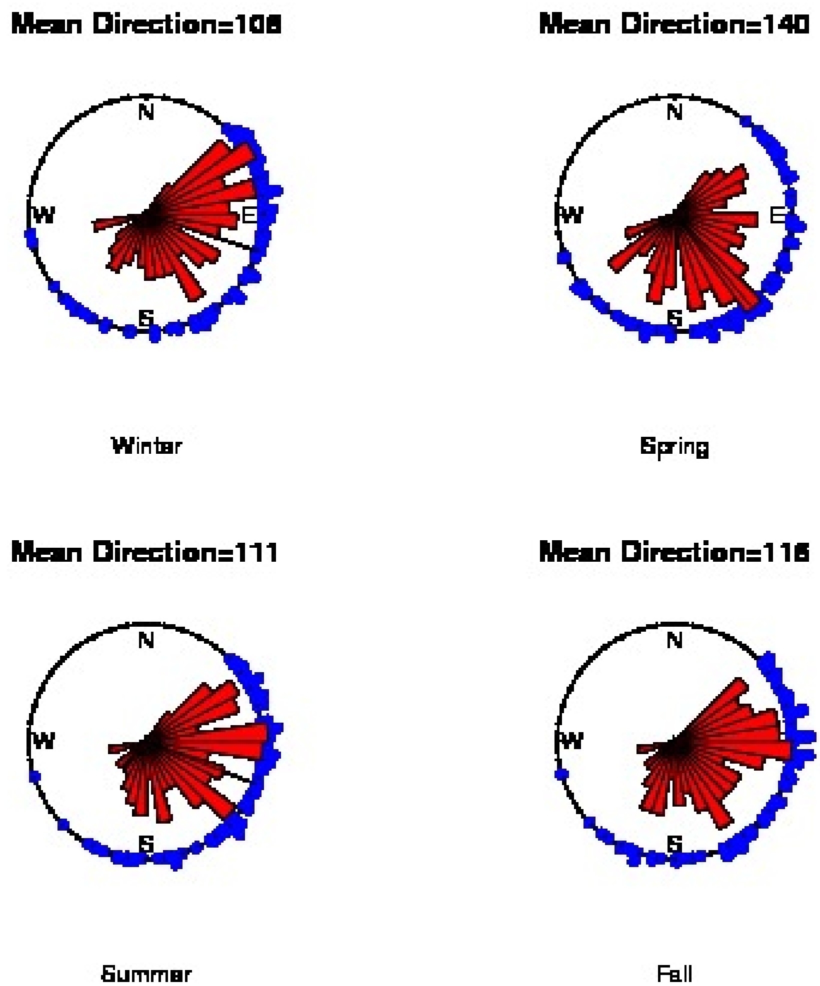


FIGURE 1. Seasonal Rose Diagrams For Ankara's Wind Data

The modified version is proposed by [11] which is based on Stephen's approximation; as suggested by [14].

After confirming the validity of the assumptions before circular ANOVA, we are now ready to run the circular ANOVA test in R. We would like to see if there is a significant difference in the mean wind directions of winter, spring, summer and fall seasons for the city of Ankara. So, we set

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4, \text{ versus } H_1\text{: At least two are distinct.}$$

The circular package in R has aov.circular option that performs circular ANOVA test. The circular ANOVA program in R has two options. First, the analysis can be performed by using F-test if the common kappa parameter (κ) is greater than 1 (if the κ parameter is between 1 and 2, then a modified F test must be performed). The second option

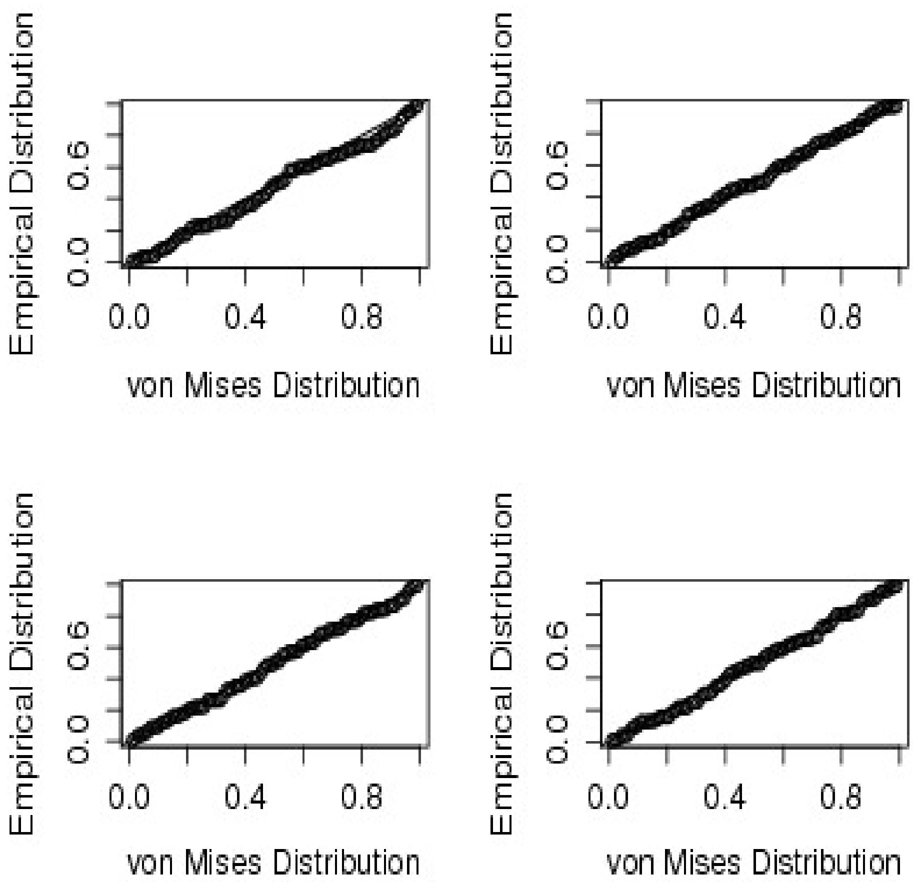


Figure 2. von Mises QQ plots of Wind Directions from Winter, Spring, Summer and Fall

performs Likelihood Ratio Test if the common kappa, κ, parameter is less than 1. Since the estimated common concentration parameter, $\hat{\kappa} = 1.69$, a modified F test is used in circular ANOVA. The result of the circular ANOVA is shown below in Table 4. Table 4

Table 4. Test of Circular ANOVA using R

Source	df	SumSquare	MeanSquare	F	Pvalue
Between	3	5.5446	1.8487	6.516	0.000266
Within	356	123.371	0.3465		
Total	359	128.917	0.3591		

implies that H_0 is rejected and, therefore, there is a significant difference among the seasonal winds directions of Ankara since the p-value of the test is 0.000266. This means that there was a seasonal difference among four seasons for the year of 2010. Visual analysis of Table 1 and Figure 1 indicates that the mean wind direction of spring season is 140^0 and looks significantly different than the other three seasons. In the next step, we will perform the circular ANOVA again without the spring season data in order to see the effect of the spring season on the analysis. The results can be seen below in Table 5. It appears that there is no significant difference among three seasons (winter, summer,

Table 5. Test of Circular ANOVA without Spring data

Source	df	SumSquare	MeanSquare	F	Pvalue
Between	2	0.3238	0.1619	0.575	0.5634
Within	267	90.8774	0.3404		
Total	269	91.2012	0.3390		

and fall) since the p-value is 0.5634. This means that the spring season has significantly different mean wind direction for the months of March, April and May. Figure 1 shows the circular plots and rose diagrams for each season, and the mean direction for spring is significantly different at $\alpha = 0.05$. The result of the circular ANOVA and also bootstrap approach could lead to new studies related to seasonal wind directions in different parts of Turkey.

4.2. Application of Nonparametric Tests. Nonparametric tests did not need prior investigation of the circular data in order to check assumptions as in the case of circular ANOVA. So, we implemented bootstrap, uniform scores test and Rao's test of homogeneity in R using circular package. Bootstrap and uniform scores test are not available in R's circular package. Therefore, a function has been written in R for those two tests. Rao's test of homogeneity is called rao.test in R via circular package. Rao's test of homogeneity gives p-value of 0.0214 for the test of H_0 which assumes all seasonal mean directions are equal. So, Rao's test implies that there is a significant difference in the seasonal wind directions of Ankara. Similar to the circular ANOVA, spring wind directions are excluded and Rao's test is applied again using winter, summer and fall data. The result shows that Rao's test gives a p-value of 0.6216 which implies no significant difference in the remaining seasons. When we run the uniform scores test on Ankara's seasonal wind data, it gives a p-value of 0.0014 which implies significant difference among the seasonal wind directions. If we repeat the test without spring season, then uniform scores test gives a p-value of 0.64 which is not significant or no difference in the mean wind directions. Bootstrap test finds a p-value of 0.0005 which is very significant and implies a difference in the seasonal mean wind directions of Ankara. If we remove the spring season from the data and run the bootstrap test again, we obtain a p-value of 0.6055. Therefore, we note that circular ANOVA and alternative nonparametric tests confirm each other and reach the same decision for Ankara's seasonal wind data.

5. Small Sample Example

5.1. Application of Circular ANOVA. Circular ANOVA and nonparametric alternatives are demonstrated under a small sample example (all samples are less than 25). The example consists of seasonal wind directions of Gorleston, England from [11]. The data have winter, spring, summer and fall wind directions, which are collected between 11:00 and 12:00AM on Sundays in 1968. Descriptive Statistics for the data shown below in Table 7. The main focus is again "is there any significant seasonal difference in the wind directions?". For this purpose, we again set $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs H_1: At least two are distinct. [11] also investigated this example and assumed that the concentration parameters of the seasonal winds are equal. [11] estimated the concentration parameter

Table 6. Descriptive Statistics for Seasonal Wind Directions in Gorleston, England

Parameters	Winter	Spring	Summer	Fall
Sample Size	12	12	13	12
Mean Direction(degrees)	272	330	57	232
Mean Resultant Length	0.4265	0.1776	0.2975	0.2656
Circ. Variance	0.5735	0.8224	0.7025	0.7344
Circ. Std. Deviation	1.3054	1.8589	1.5570	1.6282
Median Direction(degrees)	288	360	30	255.6

Table 7. Likelihood Ratio Test of Homogeneity of Seasonal Wind Directions in Gorleston, England

df	ChiSq	Pvalue
3	3.459	0.3261

from the combined samples and found it as $\hat{\kappa} = 0.24$. Moreover, it is true that all $\hat{\kappa_i} < 1$ for $i = 1, .., r$. Therefore, [11] suggests Likelihood Ratio Test (LRT) type test statistics for this problem because of too small (less than 1) concentration parameter estimate. See example 6.11 on page 165 of Mardia ([11]). Using "aov.circular" (with LRT option) in R, we find the following results: The chi-square critical value for df=3, and $\alpha = 0.05$ is 7.81 from a chi-square table. The p-value of the test is 0.3261. Thus, the result from LRT test option concludes that the seasonal wind directions are not significantly different at $\alpha = 0.05$.

5.2. Application of Nonparametric Tests. Nonparametric tests from Section 3 is executed in R to get the significance probability of the tests (p-values). In fact, [6] made a remark that the summer seasonal directions for Gorleston data appear to be different that the rest of the data and excluded it from his application of Gorleston data. Similarly, [11] used the same data set to run the uniform scores test(Mardia-Watson-Wheeler test) to investigate the homogeneity of population distributions and found that uniform score test rejects H_0 with a p-value of 0.0409. So, uniform scores test finds significant difference among seasonal wind directions. On the other hand, Rao's test of homogeneity finds a p-value of 0.9095 which does not reject H_0 that claims all mean directions are equal. One explanation of this difference in Rao's test is that it requires large samples in order to reach the nominal type-I error rate as seen in Section 6. So, as indicated by [6] and [11], the uniform score test was able to identify the significance of seasonal wind directions for Gorleston, England. Finally, bootstrap test obtains a p-value of 0.2045 for H_0 and it implies no significance difference among the seasonal wind directions.

6. Power Study

Performance of nonparametric tests are compared against the circular ANOVA by a power simulation study. Three different distribution models are considered: von Mises (ideal case for Circular ANOVA test), wrapped Cauchy and mixed von Mises with rate of mixtures of 90% and 70%, respectively. Mixed von Mises is analogues the contaminated normal distribution which is commonly used in traditional statistics to investigate data models with contaminations or outliers. We assumed that there are four random samples (for example, wind directions in four seasons) and the equality of the mean directions of four populations is the null hypothesis. So, we consider the following alternative hypothesis in order to compute the power of circular ANOVA against the nonparametric test:

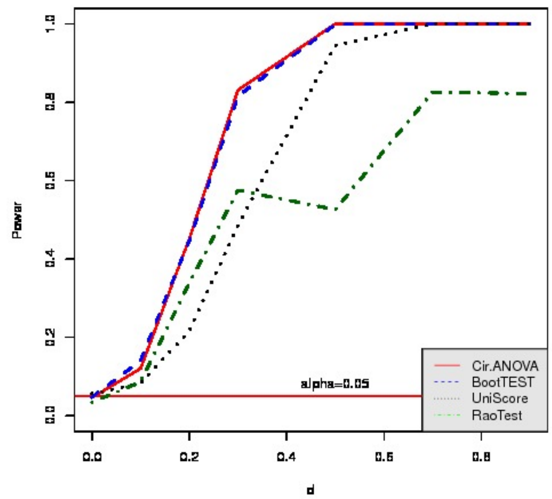
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_1 : \mu_1 + d = \mu_2,\ \mu_1 + 2d = \mu_3,\ \mu_1 + 3d = \mu_4$$

where d is a constant (shift value) that controls the alternative hypothesis. If d=0, then $H_0 = H_1$ and the tests compared in this study should reach nominal value of type-I error rate (which is set to $\alpha = 0.05$). First, Monte Carlo simulation is performed (B=1000 replications) by generating four independent random samples ($n_1 = n_2 = n_3 = n_4 = 25$) from von Mises distribution with parameters $\mu = \pi$ and $\kappa = 2$. Monte Carlo simulation finds the number of times the tests rejects H_0 under the assumption that H_1 is true for each

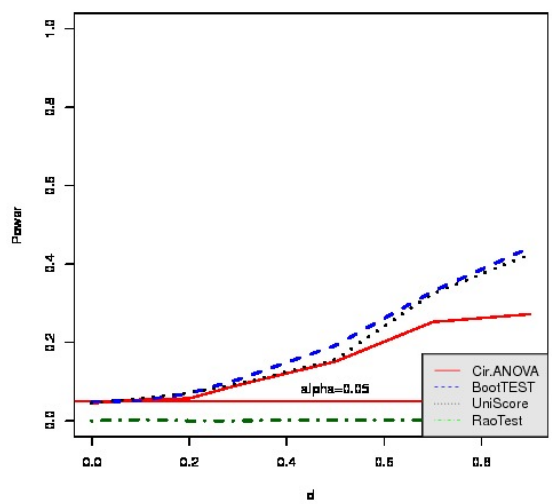
d=(0,0.1,0.2,0.3,0.5,0.7,0.9)

Then, the result is divided by B (number of replications) to find an estimate of the power. The result can be converted to the percentage that gives the empirical power of the test. Figure 3(a) shows the power curve for circular ANOVA, Bootstrap test, Rao's test of homogeneity and uniform scores test under H_1 and $\kappa = 2$ for each d.

Figure 3(a) and Table 8 show that when d=0, circular ANOVA, bootstrap and uniform scores tests have comparable estimated type-I error rates which are close to the nominal value of 0.05. On the other hand, Rao's test did not reach the nominal value of type-I error. Moreover, circular ANOVA is known to be powerful according to [6] when κ has 2 or higher and the data come from von Mises distribution. Bootstrap and Uniform score tests also worked as good as circular ANOVA under the data model and parameter assumptions. For larger shifts in the mean directions of the populations (for larger d values), uniform score test and Rao's tests started to lose some power as shown by Figure 3(a). In the next simulation, we assumed that all four samples are coming from von Mises populations and the common concentration parameter of $\kappa = 0.5$.



(a)



(b)

FIGURE 3. Circular ANOVA and nonparametric test alternatives are compared in terms of their power curves. All four samples are generated from von Mises with $\kappa = 2$ (figure a) and $\kappa = 0.5$ (figure b) parameters for each d.

Figure 3(b) shows that, when $\kappa = 0.5$, bootstrap test performed the best among the compared methods. Bootstrap test has an estimated type-I error rate of 0.049 which is very close to the nominal value of $\alpha = 0.05$. Uniform score and circular ANOVA tests are comparable at d=0 but circular ANOVA loses power at the larger shift values under H_1. As pointed out by [6], circular ANOVA requires κ parameter to be larger than 2 in order to maintain type-I error rate and its power. On the other hand, Rao's test of homogeneity did not perform well against the other three methods and did not reach the desired level of α or power. One reason could be that Rao's test requires large sample sizes to reach nominal value of type-I error. Table 8 has the numerical values of the simulations for $\kappa = 2$ and $\kappa = 0.5$ assumptions.

TABLE 8. Power simulation results for Circular ANOVA, Rao's test, Uniform Score test. All four samples are from von Mises with $\kappa = 2$ (left table) and $\kappa = 0.5$ (right table) parameters

d	CirANOVA	Boot	Uniform	Rao	CirANOVA	Boot	Uniform	Rao
d=0	0.046	0.047	0.056	0.034	0.045	0.049	0.045	0.000
d=0.1	0.120	0.140	0.083	0.086	0.054	0.051	0.057	0.003
d=0.2	0.451	0.447	0.216	0.338	0.057	0.070	0.072	0.000
d=0.3	0.831	0.816	0.486	0.575	0.091	0.105	0.093	0.000
d=0.5	0.999	0.999	0.945	0.526	0.151	0.192	0.156	0.002
d=0.7	1.000	1.000	1.000	0.826	0.252	0.330	0.324	0.002
d=0.9	1.000	1.000	1.000	0.822	0.272	0.442	0.425	0.003

In the next simulation, we considered small and large sample simulations to compare the performance of all four tests under wrapped Cauchy distribution assumption. First, four random samples of size 10 generated from wrapped Cauchy distribution with $\mu = \pi + d$ and $\rho = 0.9$ parameters. The reason that we considered the wrapped Cauchy distribution is to see the performance of circular ANOVA and alternative tests when the data come from non-von Mises models and also compare the tests under a small sample case. We repeated the same experiment for a large sample size ($n_1 = n_2 = n_3 = n_4 = 100$) using the same wrapped Cauchy distribution and parameters.

Figure 4(a) shows uniform score test performed better than bootstrap and circular ANOVA tests under a small sample case and wrapped Cauchy assumption. At d=0, uniform score test estimates the nominal type-I error rate with 0.049 which almost equals to the true rate of $\alpha = 0.05$. On the other hand, bootstrap, circular ANOVA and Rao's test did not maintain the nominal type-I error rate of $\alpha = 0.05$. Overall, Rao's test of homogeneity did not perform well again due to small sample sizes. Figure 4(b) shows the power curves under the large sample case where the random samples of size 100 created from the wrapped Cauchy distribution with $\mu = \pi + d$ and $\rho = 0.9$. Figure 4(b) shows all methods except circular ANOVA have maintained the nominal rate of type-I error as seen in Table 9. Rao's test homogeneity has an estimated type-I error rate of 0.044 for $\alpha = 0.05$ and it has shown its best performance when large samples sizes are considered. So, circular ANOVA did not perform very well under the assumption of wrapped Cauchy populations.

Table 9. Power simulation results for circular ANOVA, Rao's test, uniform score test. Four random samples of size 10 are from wrapped Cauchy with $\mu = \pi$ and $\rho = 0.9$ parameters (right table) and large sample case where $n_1 = n_2 = n_3 = n_4 = 100$ are again generated from wrapped Cauchy distribution with the same parameters (left table).

d	CirANOVA	Boot	Uniform	Rao	CirANOVA	Boot	Uniform	Rao
d=0	0.008	0.027	0.049	0.026	0.004	0.051	0.048	0.044
d=0.1	0.148	0.288	0.774	0.468	0.998	1.000	1.000	1.000
d=0.2	0.678	0.826	0.998	0.880	1.000	1.000	1.000	1.000
d=0.3	0.938	0.980	1.000	0.930	1.000	1.000	1.000	1.000
d=0.5	1.000	1.000	1.000	0.972	1.000	1.000	1.000	1.000
d=0.7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
d=0.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

In the next simulation, four independent random samples ($n_1 = n_2 = n_3 = n_4 = 25$) are generated from "mixed" von Mises distribution with proportion of the mixture is defined as $p * VonM(\mu_1 = \pi/2 + d, \kappa_1 = 3) + (1 - p)VonM(\mu_2 = 2\pi, \kappa_2 = 0.5)$ where p shows the proportion of the mixture. We will consider p=0.90 (90%-10% mixture) and p=0.70 (70%-30% mixture) proportions respectively. These model assumptions can also be considered an equivalent of contaminated normal distribution in the traditional sense. The goal is to see the performance of circular ANOVA and nonparametric tests under these assumptions that random samples come from mixture of von Mises distributions. This approach is clearly a violation of the assumption for circular ANOVA since the test requires all r populations should come from von Mises distributions with equal parameters. For each d, Monte Carlo simulation is performed and estimated power curve of each method is presented in Figure 5.

Simulation results are also shown by Table 10 below. As it can be seen from Figure 5(a) and also from Table 10, uniform scores test performed the best overall when $p = 0.90$ (90%-10% mixture). At d=0, the nominal type-I error rate (α) should be reached if a test works as expected but only uniform scores comes close to the nominal value of $\alpha = 0.05$ with estimates of 0.044. Circular ANOVA and Rao's test estimates for $\alpha = 0.05$ were 0.028 and 0.015, respectively. It could be an indication that these two tests are very conservative in rejecting H_0. Bootstrap test is also under performing since its estimated type-I error rate is 0.034 but it is slightly better than circular ANOVA and Rao's test. If we assume $p = 0.70$ ($70\% - 30\%$ mixture of von Mises distributions) and generate four random samples from this mixed von Mises distribution, simulation results show uniform scores tests have an estimate of 0.047 for $\alpha = 0.05$. It is considerably close to the nominal value of type-I error rate and indication that the test works as expected even if the data come from mixture of von Mises distribution. On the other hand, circular ANOVA and bootstrap have estimates of 0.033 and 0.039 which are much smaller then the nominal value of $\alpha = 0.05$. Again, circular ANOVA and bootstrap test look very conservative when we assume mixture of von Mises distributions with $p = 0.70$. Similarly, Rao's test did not perform well for the mixture of von Mises distributions when $p = 0.70$. Thus, uniform scores tests should be considered a better performer under contaminations and violation of having von Mises distribution assumption.

Table 10. Power simulation results for circular ANOVA, bootstrap test, Rao's test of homogeneity, and uniform score test from the mixture of von Mises populations with proportion of the mixture is 90% (right table) and 70% (left table) respectively.

d	CirANOVA	Boot	Uniform	Rao	CirANOVA	Boot	Uniform	Rao
d=0	0.028	0.034	0.044	0.015	0.033	0.039	0.047	0.003
d=0.1	0.185	0.192	0.133	0.021	0.090	0.089	0.103	0.028
d=0.2	0.635	0.6466	0.458	0.044	0.332	0.341	0.334	0.007
d=0.3	0.941	0.947	0.829	0.438	0.725	0.729	0.720	0.061
d=0.5	0.998	0.998	0.989	0.902	0.942	0.945	0.935	0.307
d=0.7	1.000	1.000	0.998	0.993	0.991	0.992	0.990	0.645
d=0.9	1.000	1.000	1.000	0.999	1.000	0.999	1.000	0.803

7. Conclusion

The main motivation of this paper was about investigating circular ANOVA (one way analysis of variance in circular data analysis) against nonparametric alternatives such as bootstrap test, uniform scores test (Mardia-Watson-Wheeler test) and Rao's test of homogeneity in the analysis of multi-sample circular data problems. Circular ANOVA requires certain assumptions as we discussed in Section 2. On the other hand, bootstrap, uniform scores, and Rao's tests are considered nonparametric tests, and they do not depend on any population distributions (see Section 3) or equality of parameters. There is also a lack of study in the literature about the comparison of circular ANOVA with alternative methods if the assumptions of circular ANOVA do not meet. So, real life examples and power analysis are performed on circular ANOVA, bootstrap, uniform scores test and Rao's test of homogeneity to observe their comparative performance under von Mises, mixed von Mises and wrapped Cauchy distribution assumptions.

Section 6 presents power simulation study which is performed to see the performance of nonparametric tests against circular ANOVA under von Mises distribution. As seen in Figure 3(a) that it is an ideal case for circular ANOVA since the test gives its best performance if $\kappa = 2$ or higher but circular ANOVA starts under performing compare to the uniform score test if $\kappa < 1$ as shown by Figure 3(b) and Table 8. Moreover, Figure 4 shows power curves of all four tests under a small and large sample cases. As we see in Figure 4(a) that uniform score test performs better than bootstrap and circular ANOVA when sample sizes are small and come from wrapped Cauchy populations. Rao's test can not compete with them if the sample sizes are too small. Next, we considered a large sample case where all four random samples have a size of 100 and the results are presented by Figure 4(b) and Table 9. As we see that all four tests have converged power curves but only bootstrap and uniform score tests have maintained the nominal type-I error rate of 0.05 which is an indication that under a large sample case bootstrap and uniform score test works as expected and detect shifts in the mean directions better than circular ANOVA. Figure 5 and Table 10 are obtained by generating four random samples (sizes of 25) from mixed von Mises with ($\mu_1 = \pi/2$, $\kappa_1 = 3$) and ($\mu_2 = 2\pi, \kappa_2 = 0.5$) with a mixture rate of $p = 0.90$ and 0.70 respectively. Figure 5(a) (also Table 10) shows that only uniform scores test is almost equal to the nominal type-I error rate of 0.05. Therefore, uniform scores test could be used without sacrificing the power of the test compare to the circular ANOVA, bootstrap and Rao's test under the mixture of von Mises distributions with $p = 0.90$. Figure 5(b) also shows uniform scores test is almost equal to the nominal type-I error rate when we assume mixed von Mises with a mixture rate of $p = 0.70$. In both cases of mixed von Mises distributions, circular ANOVA and bootstrap tests are less likely to reject H_0 when it is false since their estimates of nominal type-I rate are much smaller than $\alpha = 0.05$. Similarly, Rao's test is also under performing when we assume mixture of von Mises distributions.

We can conclude that circular ANOVA shows superiority if the data come from von Mises distribution with a common concentration parameter of $\kappa = 2$ or higher which is considered an ideal case for circular ANOVA. If $\kappa < 1$, bootstrap and uniform scores tests performs slightly better overall. If we assume mixed von Mises and wrapped Cauchy distributions, uniform scores tests performs better than circular ANOVA, bootstrap and Rao's test of homogeneity in which Rao's test requires large sample sizes in order to reach the performance of the alternative tests.

8. Acknowledgement

We greatly appreciate the comments and suggestions made by the referees and the editor of the journal in preparation and publication of this manuscript.

References

[1] Agostinelli, C. and Ulric, L. (2009). Circular Package in R, http://cran.r-project.org/web/packages/circular/circular.pdf.

[2] Batschelet, E. (1981). *Circular Statistics in Biology*, Academic Press, London.

[3] Efron, B. (1979). Bootstrap methods: Another look at the Jackknife, *Ann. Statist.*, 7,pp 1-26.

[4] Efron, B. and Tibshirani, R.J. (1993).*An introduction to the Bootstrap*, Chapman and Hall, New York.

[5] Fisher, N.I. and Hall, P. (1989). Bootstrap Confidence Regions for directional data. *Journal of the American Statistical Association*, 84, 408, pp. 996-1002.

[6] Fisher, N. I. (1993). *Statistical Analysis of Circular Data*, Cambridge University Press, New York, USA.

[7] Hall, P. (1988). On the Bootstrap and Symmetric Confidence Intervals. *Journal of the Royal Statistical Socieity. Ser. B*, 50, pp. 35-45.

[8] Harrison, D., Kanji, G.K. and Gadsden, R.J.(1986). Analysis of variance for circular data, *Journal of Applied Statistics*, vol 13,issue 2.

[9] Jammalamadaka, S. R. and Sen Gupta, A. (2001). *Topics in Circular Statistics*, World Scientific Publishing Co. Pte. Ltd, London, England.

[10] Mardia, K. V. (1967). A nonparametric test for the bivariate two-sample location problem,*Jour. of Roy. Statis. Soc. Ser. B*, B29, 320-342.

[11] Mardia, K. V. (1972). *Statistics of Directional Data*, Academic Press, London and New York.

[12] Okamura, H. and Takasuka, A. (2012). A bootstrap method for testing equality of peak months. *Population Ecology* 54,1, 169-176.

[13] Rao, J.S. (1967). Large sample tests for the homogeneity of angular data, *Sankhya*, Ser, B., 28.

[14] Stephens, M. (1969). Tests for the von Mises Distribution, *Biometrika*, 56,149-160.

[15] Tasdan, F. (2013). Technical Report: R programs for Circular ANOVA and Nonparametric Alternatives, http://www.wiu.edu/users/ft100/WindRcode.pdf

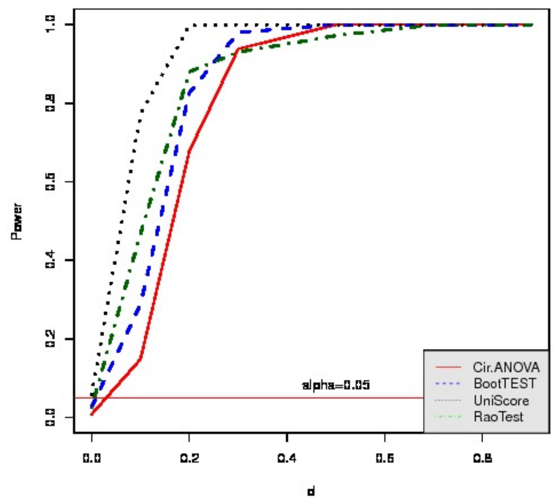
[16] Watson, G. S. (1983). *Statistics on Spheres*, Wiley, New York, USA.

[17] Watson, G. S. and Williams, E. J. (1956). On the construction of significance tests on the circle and the sphere, *Biometrika*, 43,344-352.

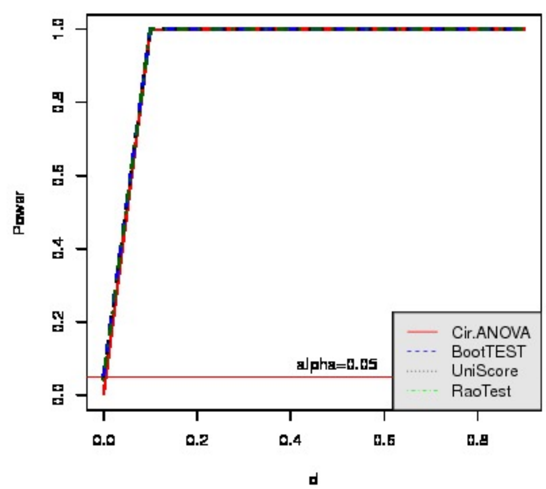
[18] Wheeler, S. and Watson, G. S. (1964). A distribution free two sample test on the circle, *Biometrika* 51, 256-7.

Appendix A

R functions that are used in this paper can be found in Tasdan ([15]). These functions require "circular" package to be installed first in order to run the functions.

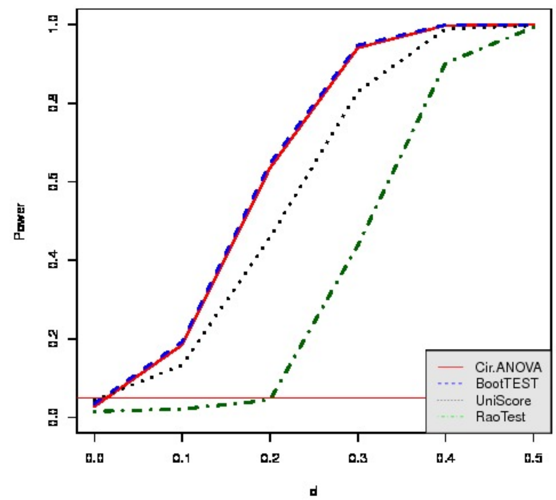


(a)

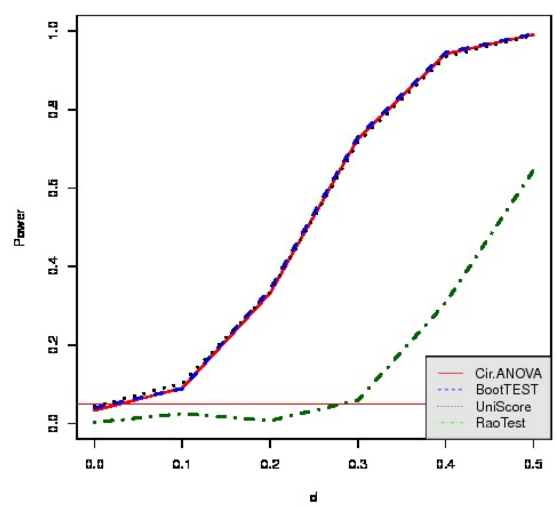


(b)

FIGURE 4. Circular ANOVA and nonparametric alternatives are compared in terms of their power curves. Figure (a) shows all four samples of size 10 (small sample case) are generated from wrapped Cauchy distribution with $\mu = \pi$ and $\rho = 0.9$ parameters and figure (b) shows large sample case where $n_1 = n_2 = n_3 = n_4 = 100$.



(a)



(b)

FIGURE 5. Circular ANOVA and nonparametric tests are compared in terms of their power curves. Figure (a) shows all four samples are generated from mixed von Mises with $\mu_1 = \pi/2 + d$, $\kappa_1 = 3$ and $\mu_2 = 2\pi$, $\kappa_2 = 0.5$ with proportion of the mixture is 90% and figure (b) shows the repeat of the simulation with proportion of 70% mixture.