

R Package: MedOr

Adriano Polpo* Carlos A. de B. Pereira†

December 24, 2012

Abstract

The purpose of this text is to provide a simple manual for the `MedOr` package for R language. In short, we give some examples on how to use the main functions of the package.

Keywords: MedOr package, median estimation, order statistics.

1 Introduction

This paper present the `MedOr` package for R ([R Core Team, 2012a](#)), for more details see [Pereira \(2012\)](#). It is an open source software under [GPLv3](#). We illustrate trough examples how to use the functions in the package.

There are two possible analysis to be performed using the package, the first one consist in a confidence interval for the median population (Section [3](#)) and the confidence statement for population median order of two or more groups (Section [4](#)).

We considered the non-parametric methods based on the order statistics using exactly distributions, we do not use any asymptotic result in the development of the procedures. Of course that some numerical approximation are used when necessary. However, they are not based on asymptotic distribution related to the sample.

2 Installing the package

In case you have not yet done so, the first thing to do before using the functions is to install and load the library. The package can be obtained from <http://code.google.com/p/medor/>.

*Federal University of São Carlos; email: polpo@ufscar.br

†University of São Paulo; email: cpereira@ime.usp.br

To install the package from the source file we use

```
> install.packages("MedOr_VERSION.tar.gz",  
+                   repos=NULL, type="source")
```

For more details in installing R or R packages see [R Core Team \(2012b\)](#). After the package has been installed, to use their functions we should “call” it by the command

```
> library("MedOr")
```

3 Confidence interval

Let X_1, \dots, X_n independent identically distributed continuous random variables (or a random sample). Our interest consist in evaluate a confidence interval based in the order statistics for the population median, with level $\alpha \in (0, 1)$.

We have that the interval $X_{(j_1)} < M_X < X_{(j_2)}$ has significance level α given by $\alpha = \Pr(B \geq j_1) - \Pr(B \geq j_2)$, where B is a random variable with binomial distribution of parameters n and $1/2$, $j_1, j_2 = 1, \dots, n$, and $j_1 < j_2$.

Since B has discrete distribution, it may not possible to find an interval with confidence level exactly equal to α (fixed in advance). In this case, we evaluate two confidence intervals for population median: 1) the interval with confidence α^L , where α^L is the biggest possible value lower (or equal) than the desired level α ; 2) the interval with confidence α^B , where α^B is the lowest possible value bigger (or equal) than the desired level α . In this case the user can choose which interval fits better to his purposes.

Example 1 In this example we simulated a random sample of size 20 from a normal distribution with mean 0 and variance 1, then we evaluate the confidence interval for the population median, which is known and equal to the mean.

```
> x <- rnorm(20, 0, 1)  
> conf.interval(x, alpha=0.95)  
  
conf.interval(x = x, alpha = 0.95)
```

Interval for population median.

Interval with confidence level lower than alpha:

confidence level: 0.9423

confidence interval: (-2.3457, 0.0645)

distance to alpha: 0.0077

interval amplitude: 2.4102

Interval with confidence level bigger than alpha:

confidence level: 0.9586

confidence interval: (-0.8372, 0.2774)

distance to alpha: 0.0086

interval amplitude: 1.1146

Total time spent: 0.00 min

4 Confidence statement

Consider that $X_j = \{X_{j1}, \dots, X_{jn_j}\}$, $j = 1, \dots, g$, are independent identically distributed continuous random variables, and X_1, \dots, X_g is an arrangement of size g , that is, we have a sample for each one of the g different groups with size n_j . Our interest consists in evaluating the confidence level for the statement $M_1 < M_2 < \dots < M_g$, the population median ordered.

Example 2 Pre-operative Gleason score provides valuable prognosis in cases of prostate cancer, in general. However, for patient's Gleason 7 it does not. This is because Gleason 7 tumors display great morphological heterogeneity among regions. The data set has the microarray data of gene RPS28 for recurrent (R) and non-recurrent (NR) Gleason 7 prostate cancer patients. The following commands evaluate the confidence statement for the population median ($M1$) of the recurrent patients to be lower than the population median ($M2$) of the non-recurrent patients.

```
> data(gleason7)
> d <- list(x1 = gleason7[1:5,1], x2 = gleason7[,2])
> conf.statement(d)

conf.statement(data = d)
```

Confidence Statement of ordered population medians:

$$M1 < 15.4155 = X1(5)$$
$$X2(5) = 15.4361 < M2$$

Confidence statement: 0.9630

Total time spent: 0.00 min

5 Remarks

This “manual” describes the basis of the `MedOr` package. We invite the user to the functions’ help pages (available with the package).

References

- R Core Team (2012a). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0.
- R Core Team (2012b). *R Installation and Administration*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-09-7.
- Pereira, C. A. de B. and Polpo, A. (2012). MedOr: Order of Medians Based on Confidence Statements. arXiv.org. <http://arxiv.org/abs/1212.5405>.