

NAME

`mannwhitney` – perform a nonparametric Mann-Whitney test on 2 sets of angular data

SYNOPSIS

`mannwhitney` [options]

DESCRIPTION

`mannwhitney` performs a nonparametric Mann-Whitney test on 2 sets of angular data. All input files and parameters are specified via command line options. The results appear on the standard output, which you may redirect to a file.

The angular data is assumed to represent cyclical values, in a range of 0 to 1, with 0 representing the start of the cycle and 1 the end. If your data are not in this format, you will need to provide the appropriate scaling. No range checking is performed on these data, so values below 0 or above 1 are allowed, and are wrapped appropriately to their proper position in the cycle.

For each set of data, the mean angle is calculated, and the samples are ranked according to their dispersion from their mean, with greater rankings indicating greater dispersion. A test statistic is calculated from the rank sums and compared against the critical value for the sample sizes at the desired probability, to determine if the dispersion is significantly different in the two sets. This test is described in section 6.10 of Edward Batschelet's book, *Circular Statistics in Biology*, 1981, Academic Press.

Options

Below is a list of the command line options allowed by `mannwhitney`. Each should be specified as a single argument, with no space between the option code and the following value.

-xfile

Specifies the input file from which the values of the first group of samples, or *data set*, will be read. This file name can be a single hyphen (-), which will cause input to be read from the standard input. The input file should contain lines of real numbers (floating point numbers in ASCII form), with one or more numbers per line, separated by spaces, tabs, commas, colons, semi-colons, or parentheses. Any number of separators can appear between columns of data, as columns are counted by the number of data values on the line. Missing data values on some lines can throw off the count. If you want a fixed field separator, and only allow one separator between columns, use the `-f` option below. Numbers will be read only from the column you specify as the X input column, using the `-xcorno` option below. Blank lines are skipped, but any other lines that don't begin with a number will cause reading to end there, unless it is among those you skip using the `-xlskip` option. Files generated by `joinnum(1)` are ideal for use as input files here, but they can come from any other source as well.

-yfile

Specifies the input file from which the second data set will be read, as for the first file above. You can use the same file name for both X and Y, if you take different columns for each data set. However, the standard input (-) should only be specified for one of these, as it can't reopen it to reread it.

If either this option or the `-xfile` option is omitted, the missing data will be made up from an ascending sequence of 1 to *n*, where *n* is the number of points taken from the other file. If you make use of this feature, it is expected that you will use an appropriate scaling factor for the sequence, to scale it to the somewhere within the 0 to 1 range. This allows you to compare one data set to an even distribution. At least one file must be specified. Usually this is the Y file.

-xlskip or -ylskip

Specifies the number of non-blank lines to skip at the start of the X or Y data file, before reading in numbers. By default, none are skipped.

-xnlines or -ynlines

Specifies the number of values to read from the X or Y data file. By default, it continues until end of file is reached. If this number is specified, and less values are read, you will get a warning

message.

-xcolno or -ycolno

Specifies the column from which numbers will be read in the X or Y data file. This is column 1 by default.

-xsscale or -ysscale

Specifies a scaling factor by which numbers read in for the X or Y data file will be multiplied. This factor is 1 by default. The angular data is assumed to represent cyclical values, in a range of 0 to 1, with 0 representing the start of the cycle and 1 the end. If your data are not in this format, you will need to provide the appropriate scaling, e.g. **.01** for data that are in a range of 0 to 100 (percentages of cycle), or **.002777777777** for data that are in a range of 0 to 360 (angle in cycle, in degrees).

-xoffset or -yoffset

Specifies an offset by which numbers read in for the X or Y data file will be shifted. This number is added to the data after it is scaled by the scaling factor. This offset is 0 by default.

-f[separator]

Specifies that data in the data files have fixed field separators, and are not free-format. If a *separator* is given, this will be the only field separator allowed in the files. If omitted, any single one of the following will be allowed as a separator: tab, comma, colon, semi-colon, or parentheses. Only one separator will be allowed between each column, so if two separators appear in a row with no number in between for the column you selected, the data value is assumed to be missing and the line is skipped.

-homing-angle

Specifies the homing angle to be used for the homing strength test. This is an alternate form of the Mann-Whitney test for angular statistics, where instead of normalizing data to their mean angles to compare dispersion in two populations, data are normalized to a specified homing angle to compare homing strength in two populations. This test is described in section 6.10.2 of Batschelet's book.

-pprobability

Specifies the probability desired for selecting the Mann-Whitney critical value. A default value of .005 is assumed. The program can only calculate critical values for the following priorities: .005, .025, .050, .950, .975, and .995. The one closest to the value you specify is the one that will be used. Depending on the sample sizes, the critical value for this probability may be calculated exactly, or approximated. If either sample size is over 20, or the combined sample size is over 32, the critical value is calculated by an improved normal approximation described by Hodges, Ramsey and Wechsler, 1990. These approximations may be slightly inaccurate, especially if one of the sample sizes is still 10 or less, and should be manually checked against a Mann-Whitney critical value table. For smaller sample sizes, where an exact probability calculation can be performed in a reasonable length of time, the critical value is determined by the method described by R. C. Milton, J. Am. Stat. Assn. 59(37): 925-934, Sept. 1964.

-v Specifies verbose output. The mean angles or homing angle will be shown, as well as a note about approximations of the critical value for small sample sizes. With two or more **-v** options, more verbosity is selected, and the program will show the normalized angles and their rankings.

SEE ALSO

joinnum(1), genplot(1)

Circular Statistics in Biology, E. Batschelet, 1981, Academic Press.