**Statistical Monitoring Program Instructions.**

**4 a. Digit Preference: Integers**

The function **integer\_check** can be used on continuous variables check if values are being given as integers (i.e. possibly rounded).

Parameters to give the function:

1. Data:

This should be in the form of a data frame with the site number in the first column (must be numeric – if a string is given please recode as numeric), the date that the measurement was taken in the second column (in the form dd/mm/yyyy) and the measurements to check in the following columns (i.e. columns 3+). If measurements were taken at different dates please split the data into more than one data frame and run the function on each data frame.

Data frames can be read in with the following code:

**options(stringsAsFactors = FALSE)**

**reg.data<-data.frame(read.table("STUDY12\_REG.txt", row.names=NULL, header=TRUE, sep="\t"))**

(This would read in a text file called *STUDY12\_REG.txt* and store it in the data frame *reg.data*.)

1. Minimum

This is the minimum number of observations a site can have and still be tested.

1. The trial name

The name of the trial. This will be used to label the output files. For example:

**trial.name<- “STUDY12”**

**Calling the function**

Once the program and the parameters above are stored in R’s memory the program can be run using the following command:

**integers\_check(data, min, trial.name)**

Where each parameter is stored as in 1-3

**The output:**

The program outputs a text file and a series of plots.

Each plotting display contains plots for up to 9 sites (each variable/measurement tested will have its own set of plots). If all of the values are non-integers then the points will fall on the line y=x. Any value that is an integer will appear as a horizontal line, i.e. it will not take a step up from the previous point. For example site 64 had 2 non-integer values which can be seen as steps and in site 68 all but the first observation are integers.

Sites with large numbers of integers (in a variable which should not be rounded) could be checked.

The plot above has the name: *“Integers\_Study12\_PREWBC\_2013-09-02\_sites\_60-71.txt”*

Where Study12 is the name of the trial (this is whatever has been specified as the trial name when calling the function) and the date would be replaced with the date the program was run. PREWBC refers to the name of the variable in this case the pre chemo WBC count. The program will take the name from the column name in the data frame specified and use this to label the plots. The Site numbers plotted are also given in the title (sites are plotted in numerical order).

The text file has a name in the form: *“Integers\_variables\_tested\_TRIALNAME\_2013-09-02.txt”*and contains information about the number of sites checked and the total number of sites for each variable

**Warnings:**

There are no error messages coded into the function. If data is not read in as above, the function may not work as it should, or possibly at all. Please take care when creating the parameters from your data.

Please ensure that the variables you are checking should be checked, only use on measurements which should not be rounded.