**Statistical Monitoring Program Instructions.**

**9. Variance check**

The function **variance\_check** can be used on repeated measures of continuous variables to pick out participants with unusually high or low variance.

Parameters to give the function:

1. all.data:

This should be in the form of a data frame with the id in the first column, the site number in the second column (both must be numeric – if a string is given please recode as numeric) followed by the continuous measurements to check (i.e. columns 3+).

Data frames can be read in with the following code:

**options(stringsAsFactors = FALSE)**

**treat.data<-data.frame(read.table("ABC02\_treat.txt", row.names=NULL, header=TRUE, sep="\t"))**

(This would read in a text file called *ABC02\_treat.txt* and store it in the data frame *treat.data*.)

1. min

min is the minimum number of non-missing observations a participant can have and still be included. Participants who have fewer than min non-missing observations will be removed; a text file reports excluded participants. This must be at least 2 and will be reset as 2 if a lower number is entered.

1. trial.name

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

**trial.name<- “ABC02”**

1. sheet.name

The name of the CRF being checked, again this will be used to label the output files. For example:

**sheet.name<- “treat”**

1. n

Variances for each participant will be calculated, participants will be classified as extreme if their variances lie more than n SDs from the mean variance of all participants.

1. ppp

This is the number of participants to be plotted on each plot (the plots are hard to read with too many participants on each, we set ppp as 50-70). This is only the rough number per plot as data must be split in such a way that sites are not split up over more than one plot.

1. zerochange

This is logical. If set as TRUE the program will just select participants with zero change between observations, i.e. 0 variance.

**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:

**variance.check(treat.data, min,"ABC02", "TREAT", n, ppp, zerochange)**

Where each parameter is stored as in 1-7

**The output:**

**The plots**

Each variable is plotted separately (number of plots needed for each is based on number of participants and ppp). The y-axis is an index which orders the participant by site and ID (each new site is labelled and split up by a horizontal line). The x-axis shows the values of the variable. Any participant with low variance is shown in a shade of pink/red and any participant with high variance in a shade of blue.

The example below shows is for the platelet data from sites 23-33 in the ABC02 treatment data.

This file is named *“Variance\_plot\_ABC02\_TREAT\_Platelets\_2014-07-25\_sites\_23-33.wmf”*,where ABC02 was given as trial.name, TREAT as sheet.name and the program was run on the 25/06/2014.



**The text files**

Three types of text files are output.

The first (see example, right) shows any participants that were excluded. There is one of these files for each variable that is checked as exclusions are based on the number of non-missing observations rather than the number of observations in total, i.e. some participants may be included for some variables but not for others. It has a file name of the form:

“*VARIANCE\_CHECK\_EXCLUDED\_PARTICIPANTS\_ABC02\_TREAT\_Platelets\_2014-07-25.txt*”

The second type lists participants with extreme variances. Again, this a text file produced for each variable tested.

This has a file name of the form: *“VARIANCE\_CHECK\_EXTREME\_VARIANCES\_*

*ABC02\_TREAT\_Platelets\_2014-07-25.txt”*

The third type is output only once, and shows details on all of the data tested (number of variable, number of participants etc).

This has a file name of the form:

*“VARIANCE\_CHECK\_DATA\_CHECKED\_ABC02\_*

*TREAT\_2014-07-25.txt”*



**Warnings:**

With the exception of values of min given as less than 2, 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.

You may also wish to consider removing large outliers (outlying points not participants with outlying variances) from the data before running this program as they may hide smaller extreme variances.