**Statistical Monitoring Program Instructions.**

**5. Comparison of means using Chernoff faces or Star plots.**

The function **mean\_check** can be applied to several continuous variables at once to check if there are differences in the means between sites.

Parameters to give the function:

1. Data:

This should be in the form of a data frame with the participant id in the first column, the site number in the second column (both can be numeric or string) and the continuous variables to check in the following columns (i.e. columns 3+).

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. min

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

1. trial.name

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

**trial.name<- “STUDY12”**

1. option

This should be set as “face” for Chernoff face plots and as “star” for star plots.

**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:

**mean\_check(data, min, trial.name, option)**

Where each parameter is stored as in 1-4

Or

**mean\_check(reg.data, 5, “Study12”, “face”)**

Parameters do not need to be stored as in 1-4; they can be entered straight into the command, i.e. to apply face plots to each site (with a minimum of 5 patients) in the Study 12 registration data the code above could be used.

**The output:**

The program outputs a Chernoff face plot of star plot.

Right: Chernoff face plot for study 12 registration data. Left, the same data using a star plot.

Up to two text files are also output.

The first, lists all sites which have been excluded because they do not have enough participants. This has a name in the form:

“*MEAN\_CHECK\_excluded\_sites\_faces\_Study12\_chemo\_2013-11-07.txt”*

Where “faces” tells you that the program applied Chernoff faces, the trial name was “Study12\_chemo” (i.e. the chemo data for study 12) and the date the program was run was the 7th of November 2013. If all sites can be included then this is not output.

The second text file is output only if “faces” is chosen as it contains information on which variable controls which facial feature (i.e. in the screen shot shown above we see that the height of the faces is controlled by the pre chemo haemoglobin values).

This has a name in the form: *“MEAN\_CHECK\_facial\_features\_Study12\_chemo\_2013-11-07.txt”.*

There is a second function within the R code which can be used to output details of any site which looks different. The function **mean\_check\_details** needs the following parameters to be specified:

1. data:

As above.

1. site:

The site number in question, for example site 78 or site 11 look quite different in the above plot so you might wish to look at them. If site was given as a string variable then it should be specified in quotation marks i.e. “UCLH”.

1. features:

This is the data output in the text file mentioned above. It can be read back in using the following code:

**features<- data.frame(read.table("MEAN\_CHECK\_facial\_features\_Study12\_chemo\_2013-11-07.txt", row.names=NULL,header=TRUE, sep="\t" ))**

1. trial.name

As above

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

**mean\_check\_details(data, 78, features, trial.name)**

(For site 78).

This outputs a text file with details for the chosen site:



The text file has a name in the form: *“MEAN\_CHECK\_details\_site78\_study12\_2013-11-07.txt”*. It contains the site number so it can be run several times for different sites and the files will not be overridden.

**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.