**Statistical Monitoring Program Instructions.**

**10 Categorical variables checker.**

The function **cat\_check** is used to investigate categorical variables. It compares the proportion of patients in each level in a site to the proportions found in all other sites.

Parameters to give the function:

1. data

This is a data frame with the site name/number (can be string or numeric) in the first column and any categorical variables to check in the following columns.

Data frames can be read in with the following code:

**options(stringsAsFactors = FALSE)**

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

(This would read in a text file called STUDY1*2\_REG.txt* and store it in the data frame *data*.*rand*)

1. 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:

**cat\_check(data, trial.name)**

Where each parameter is stored as in 1-2

**The output:**

The program outputs two text files. The first gives the p-value for each site and the second gives details of the test, i.e. the proportion of patients in each level of the variable in each site and the proportions in the rest of the data.

Right, an example of the first text file. This has a name in the form: *“CAT\_CHECK\_results\_STUDY12\_rand\_STAGE\_2014-07-21.txt”* where trial.name was specified as STUDY12 and the date (21/07/2014) was the date the program was run.

An example of the second text file can be seen below. It is named in a similar format: “*CAT\_CHECK\_results\_detail\_STUDY12\_rand\_STAGE\_2014-07-2.txt”*



**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.

The Chi squared function within R will generate automatic warnings if there are not enough patients in a level to perform the test, these warnings will appear when the function is run, however p-values for these comparisons will not have been output (the table will instead read “Not enough observations”).