# Guide to working with F-CAP

This document explains how to use the F-CAP software and set it up so it can be used with minimal effort. [Section A](#_Requirements) describes the requirements before you can run the software and [section B](#_Preparatory_steps) describes the set-up steps to use F-CAP. [Section C](#_Fully_automatic_procedure) describes the automatic procedure of creating an F-CAP and includes an expanded description of all user settings in the R-programme. If you are not able to enable Java on your system, please refer to [Section D](#_Manual_procedure_(without) which describes the manual procedure.

The F-CAP procedure can also be cut into two parts: the SAS calculations and the Analysis part. This is useful in two cases. Firstly, if SAS is not available on the computer you wish to generate the F-CAP with. In this case you run the SAS calculations ([section E](#_Partial_procedure:_SAS)) on the SAS-enabled computer. The intermediate output can be transferred to another system (e.g. a Java-enabled one) on which the analysis and the generation of the F-CAP are done ([section F](#_Partial_procedure:_Analysis)). Secondly, if you want to analyse data coming from statistical software other than SAS. As long as the data are entered in a pre-specified structured method, F-CAP is software agnostic ([section G](#_Universal_input)).

Table of contents

[A. Requirements 2](#_Toc421476222)

[B. Preparatory steps 2](#_Toc421476223)

[C. Fully automatic procedure (requires Java) 2](#_Toc421476224)

[D. Manual procedure (without Java) 4](#_Toc421476225)

[E. Partial procedure: SAS calculations 4](#_Toc421476226)

[F. Partial procedure: Analysis 5](#_Toc421476227)

[G. Universal input 5](#_Toc421476228)

[H. Miscellaneous 6](#_Toc421476229)

## Requirements

1. Functioning version of SAS (F-CAP was created and tested with version 9.3).
   1. Availability of the *traj* procedure within SAS. It can be downloaded from the [website of Bobby L. Jones](http://www.andrew.cmu.edu/user/bjones/download.htm). Installation instructions can be found at that page.
   2. If you make use of the standardised data input format ([section G](#_Universal_input)), SAS is not required
2. Functioning version of R, a freely available open-source programming language. It can be downloaded from [this page](http://cran.r-project.org/mirrors.html). At the first run of the F-CAP R-programme, the required additional modules will be automatically installed.
3. Excel (2010 or later).
4. Optional: Java. Having this available on your system enables the automatic creation of the F-CAP graphs in Excel. It can be downloaded from [this page](http://www.java.com/en/download/manual.jsp). In case this is not available, you will have to manually copy the graph values from a text-file to the Excel-file. Please ensure that the installed version of Java is compatible with your version of R (e.g. both 64-bit).

## Preparatory steps

1. Download the F-CAP zip and extract its content to a folder of your preference. The zip-file contains 3 files: fcap\_base.R (the F-CAP R-programme), fcap\_template.xlsx (the Excel template-file) and readme.pdf (a copy of this guide).

For ease of use, it is recommended to make a new copy of the R-programme for each dataset you wish to analyse, so that the user settings are stored for future use. The R-programme points to the Excel template-file, which will remain unaltered throughout use.

1. Create a folder in which the F-CAP can be saved.
2. Save your dataset as a SAS-readable .sas7bdat file.

## Fully automatic procedure (requires Java)

1. Open the F-CAP R-programme (fcap\_base.R).
2. Adjust the user settings according to your situation.
   1. The first block of settings relate to the input and output paths.

sas\_path The full path to the SAS executable, enclosed within double-quotes which are in turn enclosed within single quotes. For SAS 9.3 this could for example be: '"C:/SASHome/SASFoundation/9.3/sas.exe"'

excel\_template The full path to the F-CAP Excel template, enclosed within double-quotes. This file is contained in the F-CAP zip (see section B.1.). Example: "C:/myfiles/fcap\_template.xlsx"

input\_directory Path to the folder containing your dataset (see section B.3.). Example: "C:/myfiles/study-A/"

input\_file Name of the dataset file (see section B.3.). Example: "expression.sas7bdat"

output\_directory Path to the folder you have designated for the output (see section B.2.). This is where the F-CAP will be saved. Example: "C:/myfiles/output/"

* 1. The second block of settings are the variables used by the *traj* procedure. For a full explanation of the variables, please refer to the [*traj* homepage](http://www.andrew.cmu.edu/user/bjones/).

id Variable name of the identifier within your dataset, e.g. the participant number. Example: "cDNA"

var Dependent outcome variables within your dataset. Example: "o1-o5"

indep Independent outcome variables within your dataset. Example: "t1-t5"

model Dependent variable distribution. One of: "CNORM", "ZIP", or "LOGIT"

Additional variables can be appended to the model variable. For example, by default the minimum and maximum values for censoring are determined automatically if a CNORM model is selected. However, one can also set these manually in the following way:

model <- "CNORM; MIN 3; MAX 42"

Please note that the dependent variable distribution (CNORM) is followed by a semi-colon, whereas the final character of the statement is not. Another example which adds time-varying covariates to a ZIP model:

model <- "ZIP; TCOV c1-c5"

* 1. The third block of settings controls the F-CAP graphs.

groups\_min Minimum number of groups (k) to test in a model. Models with less latent trajectories will not be included. Absolute minimum: 1. Maximum: 19

groups\_max Maximum number of groups (k) to test in a model. Models with more latent trajectories will not be included. Absolute maximum: 20. Minimum: 2. Please note that groups\_min may not exceed groups\_max.

poly\_order Polynomial order of the trajectories. Minimum: 0, maximum 4. All trajectories will be assigned the same polynomial order.

* 1. The last block of settings relate to the user interaction with the F-CAP creation.

mode One of "FULL", "SAS", "ANALYSIS", or "UNIVERSAL". For the automatic procedure choose "FULL".

open\_file Either TRUE or FALSE. Determines whether the resulting F-CAP Excel-file will be automatically opened once it has been created. If Java is not available and a text-file with the F-CAP values is created, both the text-file and the empty F-CAP Excel-file will be opened.

1. Run the F-CAP R-programme. Please note that the calculations in SAS may take a long time, depending on the size of the dataset and the speed of your system.
2. The F-CAP file will be automatically opened (dependent on the open\_file variable, see section C.2.d.) and can be found in the pre-designated output folder (see section B.2. and output\_directory in section C.2.a.). The name is FCAP\_timestamp.xlsx.

## Manual procedure (without Java)

1. Open the F-CAP R-programme (fcap\_base.R).
2. Adjust the user settings according to your situation. See section C.2. for a full explanation of all variables. For ease of use it is recommended to set open\_file to TRUE (see section C.2.d.).
3. Run the F-CAP R-programme. Please note that the calculations in SAS may take a long time, depending on the size of the dataset and the speed of your system.
4. Select all values in the F-CAP text-file (ctrl+A), copy them (ctrl+C), and paste them in the F-CAP Excel-file on the data-sheet in cell B3 (ctrl+V). Switch to the plots-sheet to view the F-CAP. The text-file and the Excel-file are both located in the pre-designated output folder (see section B.2. and output\_directory in section C.2.a.). The names of the files are FCAP\_timestamp.txt and FCAP\_timestamp.xlsx.

## Partial procedure: SAS calculations

1. Open the F-CAP R-programme (fcap\_base.R).
2. Adjust the user settings according to your situation. See section C.2. for a full explanation of all variables.  
   Ensure that mode is set to "SAS". You do not need to set the excel\_template variable or have the template available on your system.
3. Run the F-CAP R-programme. Please note that the calculations in SAS may take a long time, depending on the size of the dataset and the speed of your system.
4. The output consists of a series of .csv files in the output folder (see section B.2. and output\_directory in section C.2.a.). They are named fcap\_bic\_groupnumber.csv and fcap\_prob\_groupnumber.csv. The files can be used as input for the Partial procedure: Analysis (section F). This can also be carried out upon a different computer by transferring the .csv-files to another machine. The files are overwritten if the F-CAP software is run again using the same output-directory.

## Partial procedure: Analysis

1. Ensure that you have the intermediate .csv output-files (see section E.4.) available in a single folder.
2. Open the F-CAP R- programme (fcap\_base.R).
3. Adjust the user settings according to your situation. See section C.2. for a full explanation of all variables.  
   You do not need to set sas\_path, input\_directory, input\_file, the block of settings related to the *traj* procedure (id, var, indep, and model), and poly\_order.  
   Ensure that output\_directory points to the folder which contains the intermediate .csv output-files.  
   Set mode to "ANALYSIS".
4. Run the F-CAP R-programme.
5. The F-CAP file will be automatically opened (dependent on the open\_file variable, see section C.2.d.) and can be found in the pre-designated output folder (see section B.2. and output\_directory in section C.2.a.). The name is FCAP\_timestamp.xlsx. If Java is not available, please follow step 4 of section D.

## Universal input

1. Data from software sources other than SAS need to be entered in a structured format of CSV-files (comma separated values). For each model a file named universal\_#.csv needs to be created, where the # is replaced by the number of trajectories within that model. Thus, if you wish to generate an F-CAP of models ranging from 5 to 15 trajectories, ten CSV-files need to be fed into the programme.
2. The structure of the CSV-files is as follows:

AIC,BIC,L,CONVERGENCE

PROB1,PROB2,PROB3,PROBn,GROUP

The first line consists of values pertaining to the entire model. Convergence has to be one of: True, False, Singular. If a model converges without errors, True can be used, or this parameter can be skipped entirely.

The second line, and any line after that, states the probability of an entry to be assigned to a specific group, as well as the group to which the entry was actually assigned. There is no limit to the number of entries. Note that entry identifiers should not be provided. An example of valid input for a model with ten entries (e.g. ten participants in a clinical study) and three trajectories is as follows:

-923.8511692,-940.342756,-913.8511692,True

0.697914,0.061932,0.240154,1

0.940268,0.008749,0.050983,1

0.001516,0.001025,0.997459,3

0.894734,0.046908,0.058358,1

0.011071,0.967260,0.021669,2

0.095617,0.751091,0.153292,2

0.022546,0.162716,0.814738,3

0.029783,0.940407,0.029810,2

0.073581,0.855045,0.071374,2

0.038848,0.066979,0.894173,3

1. There are only a few parameters that need to be set in the user settings to analyse universal input. For full details see section C.2. The parameters that need to be set are:

excel\_template

input\_directory Path to the folder containing the CSV-files.

output\_directory

groups\_min

groups\_max

mode Set this to "UNIVERSAL".

## Miscellaneous

If the F-CAP (the Excel-file) is saved as a pdf-file, it results in a single A4-page file containing all graphs. This is an easy method to view and share the F-CAP.