# A short overview of EpiData Analysis (2018)

## EpiData Analysis is used for data management and quantitative statistical data analysis.

Use EpiData Analysis when you want to do simple or comprehensive data management and basic descriptive statistical analyses. As part of the modern EpiData suite, Analysis was designed to work with all of the features of the EpiData data file. It is available for Linux, Windows and MacOS and it can also read some other data formats, including tabular data copied from other applications. Extended statistical modelling must be done with other software such as R, Stata etc.

**EpiDataAnalysis Classic**, which was developed from 2004-2014 is still available from [www.epidata.dk](http://www.epidata.dk/). The classic version will not be further updated beyond version 2.2.

Analysis v1.0 was first released in April 2018 with the following functionality:

* Comprehensive data management, data verification, checks and documentation features. Data may be verified (validate and check data).
* Read & write several data formats (epx/epz, encrypted epx/epz, csv, Stata) and export DDI-v3.1. All character handling is UTF-8 compatible. Results may be saved in text or html format. The previous rec+chk format of Classic EpiData may be read, but not written.
* Capability to handle large data set. All functions are validated with datasets having up to 125,000 observations and 250 variables.
* Easily handles related datasets and encrypted EpiData projects with user logging.
* Initial suite of statistical analyses: count, frequencies, means.

Following the release of v1.0, additional functions will include: collapsing of data (aggregate), crosstabulation, table estimation and graphs. See [www.epidata.dk](http://www.epidata.dk/) for news and updates.

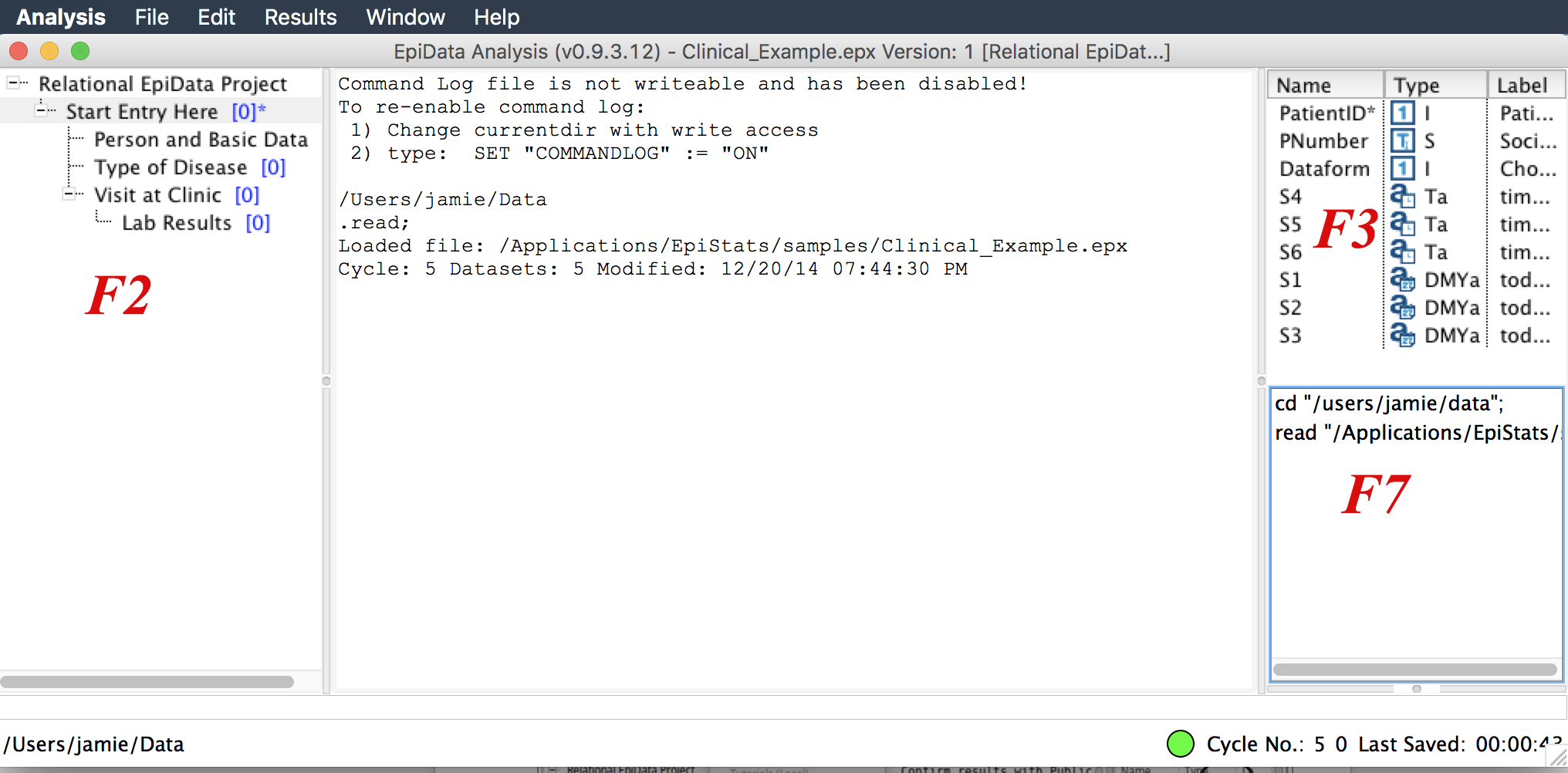
## Installation

EpiData is freely available to download from [www.epidata.dk](http://www.epidata.dk/). It can be installed easily on Linux, Windows or MacOS. For Windows, there is a combined installer for EpiData Manager, EntryClient and Analysis. Analysis will not interfere with the setup of your computer. Each of the three EpiData applications consists of a single executable file and a number of help files in html or pdf format.

## Simplicity and sophistication combined

Analysis provides a clean interface, so you can get right to work. When you start Analysis, you will see its main menu, the output window, a command line window and a status bar that has some basic information about your data set.

At present, Analysis is mostly command driven—you enter simple commands and it shows the results. This is ideally suited to quick exploratory analysis of your data or manipulation of the data prior to more sophisticated statistical or epidemiological analysis. It provides all of the control available in EpiData Entry, including data security and integrity features. It easily handles relational data that is native to all EpiData applications.

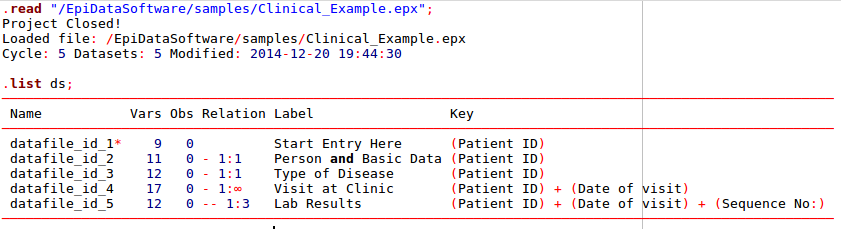


You may also choose to display more information in the left and right sidebars. After reading a file, you can see the full data structure, including related datasets (function key F2), variables (F3) and command history (F7). Other function keys will open the program editor (F5), data browser (F6) and basic help documents.

## A first analysis

EpiData comes with some sample data sets to help you become familiar with the way it works. If you are new to Analysis or want to see how it differs from Classic EpiData Analysis, follow along.

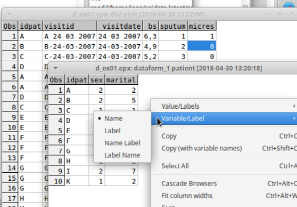
**Open (read) the data**. In the menu bar, select File, Open and find the folder/directory where you installed EpiData. You can also type *read* in the command line.

In the subfolder “samples”, Select **clinical example.epx**. Depending on your computer defaults, you may not see the extension **epx**. Double click on the file name and you will see some information about the data in the output window.

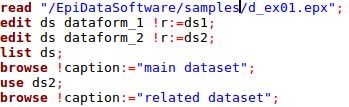
This sample project file contains a number of related data sets. The command **list ds** shows these, along with some information on each. A copy of data is read into memory, so that all data manipulation and analyses are done quickly.

**Browse the data**. Every time you open a data file it is good practice to view the data. You do this with the F6 function key or by typing *browse* in the command line. Analysis shows you a spreadsheet of the data. You can move and resize the data browser window. Keep it open while you do your analyses. Browse opens a new view of the data, exactly how they are at the time of issuing the **browse** command.

With options when you start browse you may arrange or change the view, e.g. value labels or values or changing the caption as shown below.

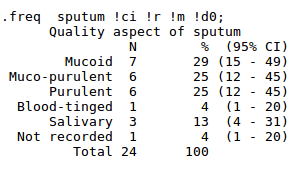
****

**Edit data, structure or content.** In analysis several commands exist, which will create, edit or modify datasets or variables. The command **edit ds** is shown here. The help file shows the possible options, variations and suggestions.



The commands show above will read a project file, rename the datasets inside the project, browse the first dataset, change to the second dataset **(use)** and show it in a browser.

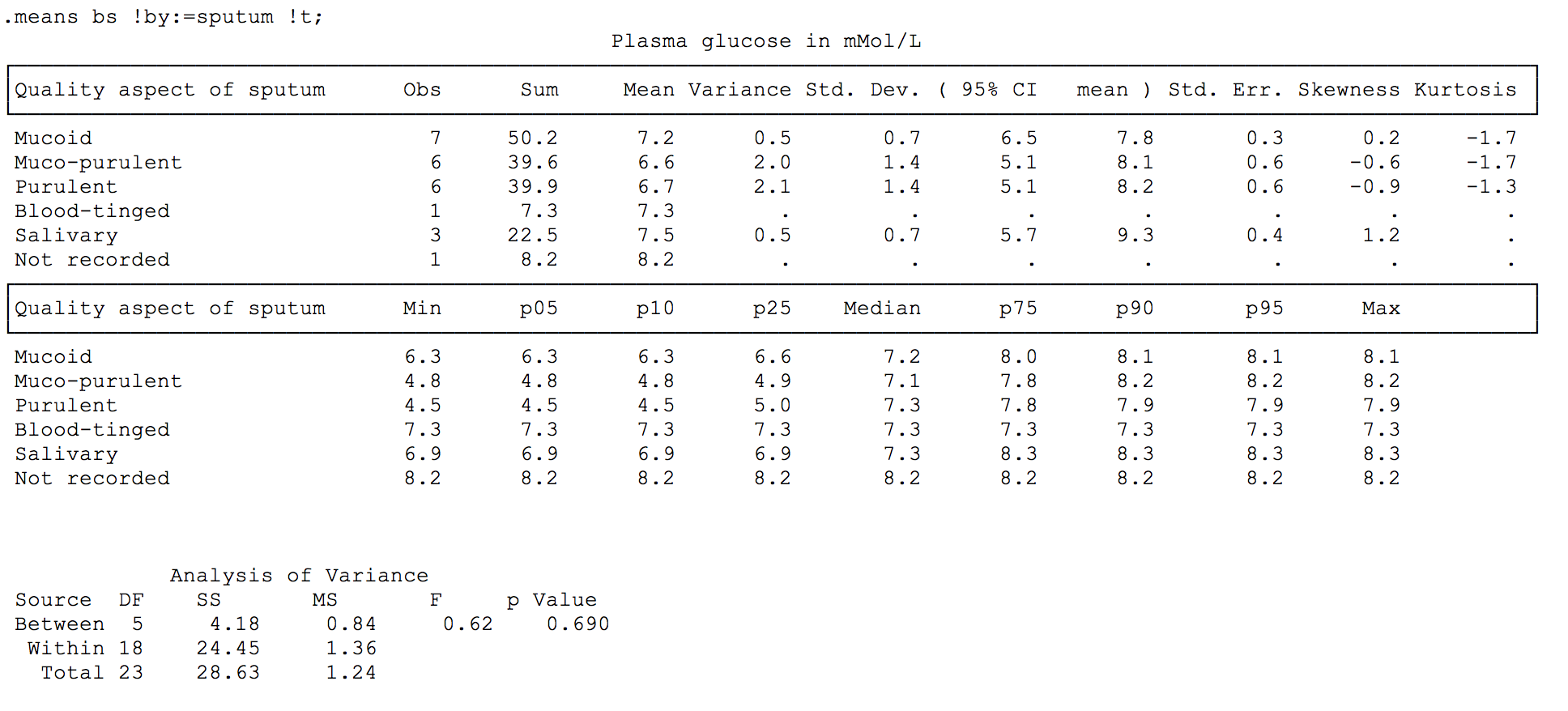
**Get a table of frequencies**. In the command prompt (F4 will move the cursor there quickly) write **freq sputum**. If you prefer to avoid typing variable names, you can open the variables window (F3) and then double click on the variable you want. You will see that the variable name is copied to the command prompt. If you now press enter, the frequency table will be shown.



In the example above, there are options shown after **freq sputum**. If you have the History (F7) window open, you can double-click on the **freq** command and it will appear in the command line. You can now add the options **!ci !r** etc.. You can also use the up and down arrows to retrieve previous commands.

## Showing means of age by time

Now type in the command line **means** and choose the variable **bs**, and hit enter. After reviewing the results press the up arrow, which will show the command again. Add to the end of the freq command: **!by:= sputum !t**

Now you see a stratified analysis of the means, with an analysis of variance.

## Get further acquainted with Analysis

Resize the program windows by dragging the sides or the separators between output window (viewer) and right side or left side windows. You can save the current layout using the **window** menu and selecting **Default windowing**.

Try to change the active folder via the file menu and notice that the command history file, **commandlog.pgm** is left behind and a new one started in the new folder. This file allows you to redo your analysis, perhaps after cleaning up the data or adding more data. You can load this file into the program editor (F5). You can modify commands and run them from within the editor, and then save your commands in a file for future use.

Try the help menu. If you are connected to internet you can click on **Check Version** to compare your version with the most up to date version at [www.epidata.dk](http://www.epidata.dk/).

## Further introduction

In a specific project try more of the data management commands and you will soon get more experience. Find inspiration in example pgm files. If you have used Classic Analysis, look at the document on how to upgrade previous pgm files, which is also installed in the local documentation folder.

You have now acquainted yourself with Analysis.

## Further help

All users are encouraged to seek information and advice from other users on the EpiData discussion list. You may join that list at: <http://lists.umanitoba.ca/mailman/listinfo/epidata-list>. By joining the list, you will receive information on updates and help us in deciding how to proceed with further development. The software comes with example files and further documentation. This is available through the help menu.

## Flowsheet

A simplified flowsheet of how EpiData Analysis is working is shown on the next page. **Blue parts** are optional, **black parts** are in memory and **red parts** save data permanently to disk. Note that Analysis always works with a copy of data. Your data on the disk are not changed unless you tell Analysis to do so. Note also that the command history file is saved when you quit Analysis.

## Flowsheet of working with EpiData Analysis

**Keep a COPY of data in memory**

**Read a data file**

**(epx, csv, dta, rec)**

For files with metadata (epx, dta). Read variable + value labels and further project information (epx), including missing value definitions.

Add new variables **new var**

Change contents of variables

**edit**

**if ... then ...,**

**select ... do**

Change sorting

**sort** command

**Modify data in memory**

**Save data to disk**

Command **save**

**Show results on the Screen**

Write results to disk:

**save ... !output**

**”Next” Analysis command**

**Stop Analysis**

quit or exit

All commands are saved in commandlog.pgm

## Improving Analysis

If you find errors or bugs when using the program or have suggestions for improvement, please discuss these on the EpiData-list.

Suggested citation of EpiData Analysis program:

(see later)

## Funding and acknowledgements.

An updated list of attained funding is available at <http://www.epidata.dk/funding.htm>. Further credits and acknowledgements at: [http://www.epidata.dk/credit.htm](http://www.epidata.dk/funding.htm) . International translations made to several languages, see [http://www.epidata.dk](http://www.epidata.dk/) For donations to further development see help file or send an e-mail to [info@epidata.dk](mailto:info@epidata.dk). Isolated parts of the source code are based on freeware and shareware components. Please consult credit pages.

## License

The program as such cannot be sold for money or service value. It is absolutely free. There can be NO charge taken by a web site for downloading of EpiData. EpiData software is distributed according to a Creative Commons license.

## Disclaimer

The EpiData Analysis software program was developed and tested to ensure the quality of analysis and documentation of data. We made every possible effort in producing a fail-safe application, but cannot, in any circumstance, be held responsible for errors, loss of data, work time or other losses incurred by or in relation to the program. As with any other software, you best protection is to back up your data.