A short overview of the new EpiData Analysis (2018)

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.

EpiData Analysis Classic, which was developed from 2004-2014 is still available from www.epidata.dk. The classic version will not be further updated beyond version 2.2.

The revised EpiData Analysis was first released as v1.0 with this functionality:

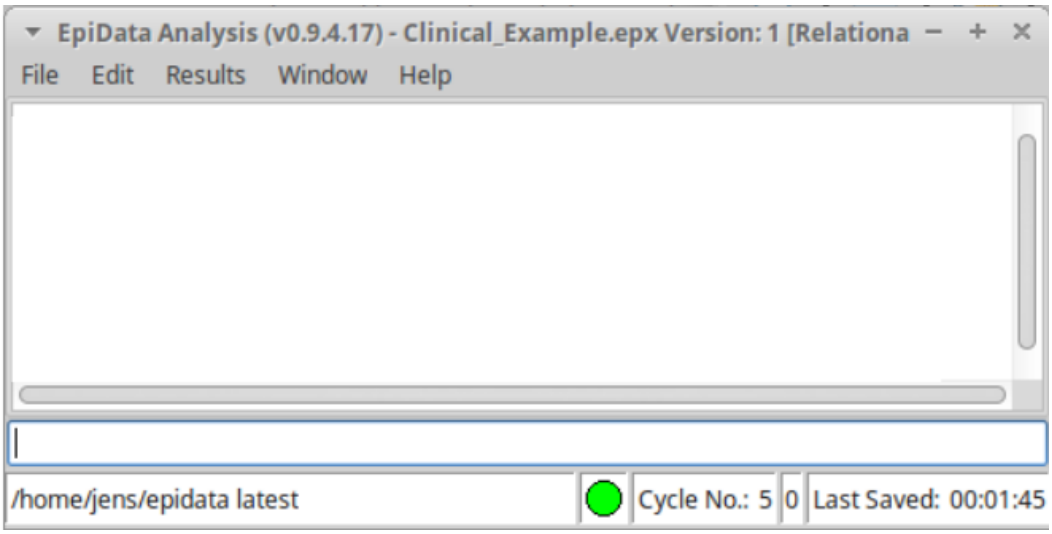
- Comprehensive data management, data verification, checks and documentation features. Data may be verified (validate and check data)
- Read & write several data formats (epx/epz, encrypted epx/epz, csv, Stata) and export DDI-v3.1. All character handling is UTF-8 compatible. Results may be saved in text or html format. The previous rec+chk format may be read, but not written.
- Analysis commands in v1.0: count, frequencies, means. Following release of v1.0 collapsing of data (aggregate), crosstables, table estimation and graphs will be implemented. See www.epidata.dk for news and updates.
- Analysis functions are validated with several batch tests, including datasets up to 125,000 observations and 250 variables. We also validate related datasets and encrypted projects with user logging.

Installation

EpiData is freely available to download from 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. EpiData software 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, and a few additional library files for encryption.

Simplicity and sophistication combined

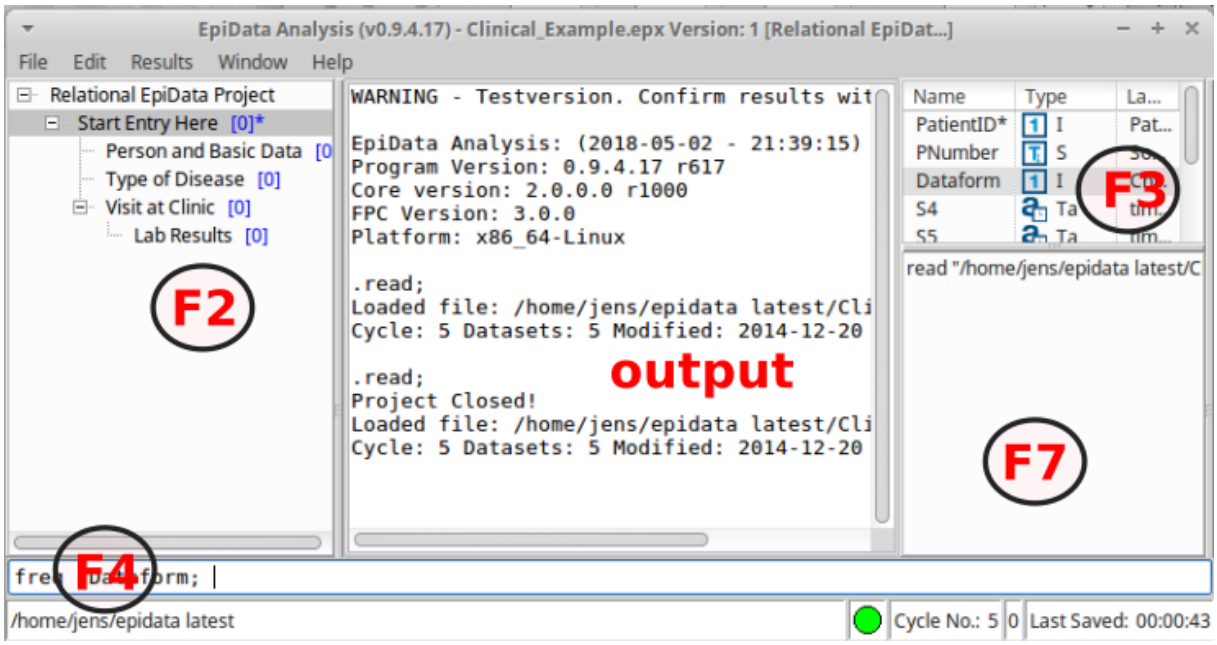
Analysis provides a clean interface, so you can get right to work. When you start Analysis, you will see its main menu, the output window showing version of the software, a command line window and a status bar that has some basic information about your data set.



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 EntryClient, including data security, encryption and integrity features. It easily handles relational data that is native to all EpiData applications.

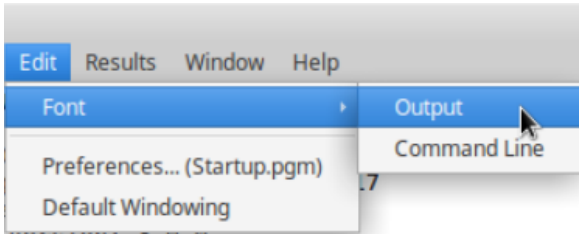
(Note: in this document, keys you press or commands you type are shown in ***bold italics***.)

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.



Flowsheet A simplified flowsheet of how EpiData Analysis is working is shown at the end of this document. Note that the Analysis always works with a copy of the data. It is good practice to use a program (.pgm file) to document changes or analysis and save updated project and program files under a new name. All commands that you type are saved to the file commandlog.pgm (one copy will appear in each active folder).

Change font size or colours. Temporary changes may be made via the **edit** in the main menu, for permanent changes every time you use Analysis open the **startup.pgm** via the **preferences**. This will open the editor, where you could easily insert all set options and see choices. An example is to change from the default text output format to html.




```
.set "OUTPUT FORMAT" := "html"; // legal values: HTML / TEXT
OUTPUT FORMAT = html (was: TEXT )
```

Your first run: Start by reading a datafile via the menu. Find **Open** in the file menu to the left on your screen and point your disk selector to where you installed the EpiData Analysis programme in the subfolder “samples” - select one of the project (.epx) files.

Open the history window (***F7***) and notice that the read command and data file name was copied here.

Here, the file **Clinical Example.epx** was read. To see the datasets within the project file issue a new command, ***list ds*** and you will see the following:

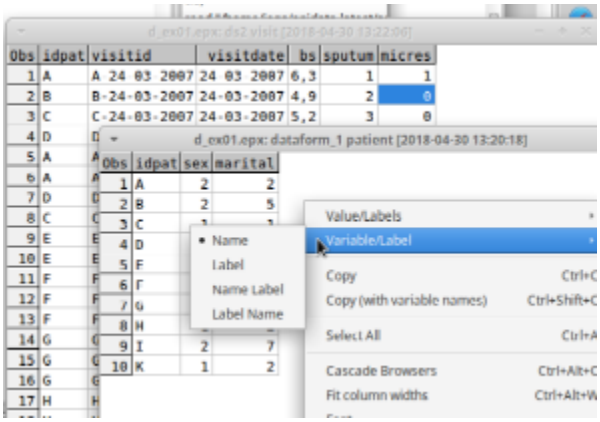
```
.read "/EpiDataSoftware/samples/Clinical_Example.epx";
Project Closed!
Loaded file: /EpiDataSoftware/samples/Clinical_Example.epx
Cycle: 5 Datasets: 5 Modified: 2014-12-20 19:44:30

.list ds;
```

Name	Vars	Obs	Relation	Label	Key
datafile_id_1*	9	0		Start Entry Here	(Patient ID)
datafile_id_2	11	0	- 1:1	Person **and** Basic Data	(Patient ID)
datafile_id_3	12	0	- 1:1	Type of Disease	(Patient ID)
datafile_id_4	17	0	- 1:∞	Visit at Clinic	(Patient ID) + (Date of visit)
datafile_id_5	12	0	-- 1:3	Lab Results	(Patient ID) + (Date of visit) + (Sequence No:)

Browse data Every time you open a data file it is good practice to view the data.

Use the browse command or press **(*F6*)** key and you will see the data in a grid. Browse opens a new copy of the data at the time of issuing the ***browse*** command.

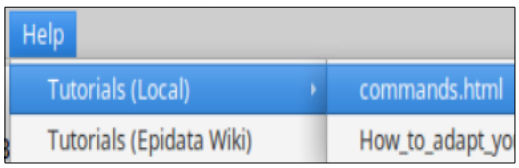
With options when you start browse you may arrange or change the view, e.g. value labels or values or changing the caption as shown below.



Edit data, structure or content.
In analysis several commands exist, which will create, edit or modify datasets or variables. Start by reading the **d_ex01.epx** file.

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

For explanation of ***edit ds*** and ***use***, consult the file **commands.html**, which you find in the help menu under **Tutorials (local)**.



Frequencies. Get a table of frequencies. In the command prompt (F4 will move the cursor there quickly) write ***freq agegrp***. If you prefer to avoid typing the variable name, 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.

```
.freq  sputum !ci !r !m !d0;
     Quality aspect of sputum
                 N        %  (95% CI)
        Mucoid   7       29 (15 - 49)
 Muco-purulent   6       25 (12 - 45)
     Purulent    6       25 (12 - 45)
  Blood-tinged   1        4  (1 - 20)
     Salivary    3       13  (4 - 31)
  Not recorded   1        4  (1 - 20)
         Total  24      100
```

How do the numbers stack up? If you add ***!r*** and ***!ci*** to the command line before pressing enter, then percentages and confidence intervals are shown to the right of the numbers of observations. With ***!m*** any missing will be included and with ***!dx*** (here x is zero, so ***!d0***) there will be no decimals on the percentages. The default is 1 decimal.

The means command has been updated from Classic Analysis. The following will guide you through use of ***means***.

First, open the example fil bromar.epx like you opened the previous project file. You may open via the file menu or by typing, ***read "bromar.epx" !c;*** The addition of ***!c*** will close the project you were previously working on. This project has data from marathon runners in Denmark. Note that you need the quotation marks around the file name.

Are male runners older or younger than females? To answer this, you will use the means command. In the command line, type ***means age***. You can always add the variable name using the variables sidebar.

```
means age;
```

age Age (1996-year of birth)

Obs	Sum	Mean	Variance	Std. Dev.	(95% CI	mean)	Std. Err.	Skewness	Kurtosis
3786	153740.0	40.6	96.3	9.8	40.3	40.9	0.2	0.1	-3.1

Min	p05	p10	p25	Median	p75	p90	p95	Max
16.0	24.0	28.0	33.0	41.0	48.0	53.0	56.0	84.0

Notice that you get a lot of information here, but we will focus on the mean age (40.6 years) of all runners.

We can compare the mean ages of men to that of women, by adding ***!by :=*** ***sex*** to our command. With the cursor in the command line (***F4***), press the ***up arrow*** once and you see that the original command is there. Add in the ***!by*** part and press ***enter*** to get your analysis of age stratified by sex.

```
.means age !by:=sex;
```

Age (1996-year of birth)

sex	Obs	Sum	Mean	Variance	Std. Dev.	(95% CI	mean)	Std. Err.	Skewness	Kurtosis
F	463	20094.0	**43.4**	78.0	8.8	42.6	44.2	0.4	-0.2	0.0
M	3323	133646.0	**40.2**	97.6	9.9	39.9	40.6	0.2	0.1	-3.1

sex	Min	p05	p10	p25	Median	p75	p90	p95	Max
F	19.0	28.0	30.0	38.0	44.0	50.0	53.0	56.0	70.0
M	16.0	24.0	27.0	33.0	40.0	47.0	53.0	56.0	84.

It looks like men runners are, on average, younger than women runners (40.2 vs 43.4 years).

Is this difference statistically significant? Go back to the command line, press ***up arrow*** and add ***!t*** to the end of the command.

Now, in addition to the results you already have seen, you get the results of an analysis of variance, which is testing whether the difference in mean age could have occurred by chance.

```
               Analysis of Variance
 Source   DF     SS        MS       F     p Value
 Between    1   4112.28  4112.28    43.20    0.000
  Within 3784 360234.47    95.20
   Total 3785 364346.75    96.26

Bartlett's Test of homogeneity of variances
 DF Chi-square p Value
  1       9.65    0.002
```

Now we see that the difference in mean age is highly significant (p<0.001). However, caution is advised because the F-test done here assumes that variances in age are equal in both groups. Analysis provides one test of this assumption, Bartlett's Test. We see that this assumption may not be met because Bartlett's test has p=0.002. So further considerations (and analysis) should manage this potential problem and other aspects. See statistics textbooks for principles.

Get further acquainted:

1. Resize the program by dragging in sides or the separator between output window (viewer) and right side parts. Save current position in window menu. **Save Window Position** or change to **default windowing** if something goes wrong.
2. Try to change the active folder via the file menu and notice that the file **commandlog.pgm** is left behind and a new one started in the new folder.
3. Run commands from within the editor and save program files for future use.
4. Try the help menu. If you are connected to internet, you may click on **Check Version**, which will compare your version with the most updated on www.epidata.dk
5. In a specific project try other data management and soon you will get more experience. Find inspiration in example program files and the document on **how to upgrade previous pgm files**, which is also installed in the local documentation folder.

Support

If you find errors, bugs or have suggestions for improvement please discuss in the EpiData-list available at http://lists.umanitoba.ca/mailman/listinfo/epidata-list

Suggested citation of EpiData Analysis program:

(to be added 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 . International translations made to several languages, see Http://www.epidata.dk For donations to further development see help file or send an e-mail to info@epidata.dk. Isolated parts of source code based on freeware and shareware components. Please consult credit pages.

Disclaimer

The EpiData Analysis software program was developed and tested to ensure fail-safe analysis and documentation of data. We made every possible effort in producing a fail-safe program, 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.

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 permanently to disk. Note that the programme always works with a copy of data. Your data on the disk are not changed unless you as a user instruct the programme to do so. Note also that the commands issued are saved when you exit.