# 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.

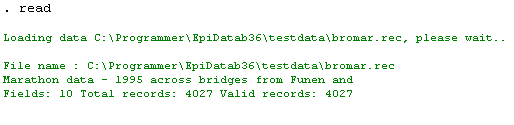
You may also choose to display more information in the left and right sidebars. After reading a file, you can see the full data structure, including related datasets (function key F2), variables (F3) and command history (F7). Other function keys will open the program editor (F5), data browser (F6) and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

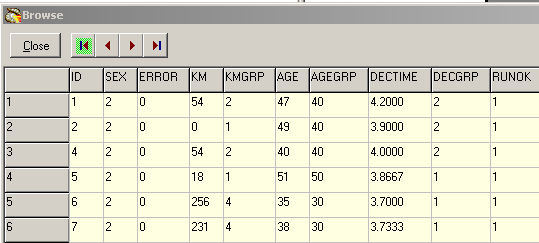
**Open (read) the data**. In the menu bar, select File, Open and find the folder/directory where you installed EpiData. You can also type *read* in the command line.



In the subfolder “samples”, Select “bromar.rec” or bromar.epx. Depending on you computer defaults, you may not see the extension .rec or .epx. Double click on the file name and you will see some information about the data in the output window. How many records and fields ?

This data set has data gathered at a Marathon race. 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.



**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 variable, 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.

**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.

.freq agegrp !r !ci;

agegrp Age (10 year intervals)

N % (95% CI)

10 35 0.9 (0.7 - 1.3)

20 489 12.9 (11.9 - 14.0)

30 1198 31.6 (30.2 - 33.1)

40 1337 35.3 (33.8 - 36.9)

50 642 17.0 (15.8 - 18.2)

60 75 2.0 (1.6 - 2.5)

70 9 0.2 (0.1 - 0.5)

80 1 0.0 (0.0 - 0.1)

Total 3786 100.0

Repeat this to get frequencies by sex.

.freq sex !r !ci;

sex

N % (95% CI)

F 490 12.2 (11.2 - 13.2)

M 3537 87.8 (86.8 - 88.8)

Total 4027 100.0

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 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 on test of this assumption, Bartlett's Test. We see that this assumption may not be met (p=0.002).

## Which elements are on the screen ?

You have now acquainted yourself with Analysis.

To get acquainted with the windows try the following:

1. Switch the extra windows on the right on and off a few times: Press keys: F2, F3 or F7

2. Resize the program by dragging in sides or the separator between output window (viewer) and right side parts.

3. Save current position in window menu. ”Save Window Position”

4. Try to change folder by clicking on the lower left side of the statusbar.

5. Try the editor. From within you can run commands or save pgm files for future use.

6. Try the help menu. If you are connected to internet you can activate ”Check Version” part which will compare your version with the most updated on on www.epidata.dk

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. Distributed according to a CC license.

## 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 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.

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**

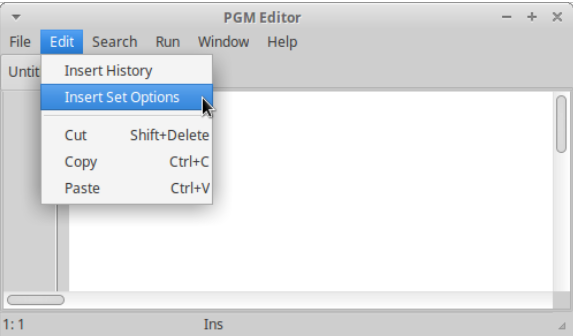
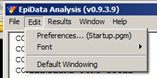
**”Next” Analysis command**

**Stop Analysis**

quit or exit

All commands are saved in commandlog.pgm

A. Start EpiData from either the programme group where you installed the program­me or an icon on Your desktop.

B. First time you run EpiData Analysis You may want to change some defaults or font sizes. Look in the main menu under “Edit” .

If you want to change these permanently then start instead the "preferences", which will open the editor. Here you may either add history or insert all possible set options.

C. You may then run (see editor menu) the sets defined or save and restart the software.

## Further introduction

Try the programme as it is and experience how you can soon get into more commands or features. If you press F1 an overview of help files is shown and if you press find (Ctrl+F) you can soon see how to get further explanations.

## Support

Sources for support:

1. Read the help files (press F1)

2. Unfortunately we do not have resources for support of questions in general. Refer these to the EpiData-list at http://lists.umanitoba.ca/mailman/listinfo/epidata-list

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](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 source code 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. Distributed according to a CC license.

## 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.