A short overview of analysis of data with

EpiData Analysis

EpiData Analysis is a program for data analysis and data management.

Use EpiData analysis when you want to do basic descriptive statistical analyses, modifications or tabulation of data. Extended analysis such as statistical modelling can be done with other software such as Stata, R etc.

EpiData Analysis is based on the same principles as EpiData Entry. If you properly define, document and verify data with EpiData Entry, the definitions are also available in EpiData Analysis. E.g. specified legal values with attached text labels(1 = No 2= Yes) or definitions of missing values. When reading data EpiData Analysis will do some control based on variable definitions, e.g. all dates are verified.

EpiData Analysis is suitable for small as well as rather large datasets. simple datasets like one questionnaire as well as datasets with many or branching dataforms.

The principle of EpiData is rooted in the simplicity of the dos program Epi Info v6, which has many users around the world. EpiData Analysis is used for: Basic descriptive Analysis of quantitive data, Defining and modifying data, Editing / correcting data already entered, Graphing Data, Asserting that the data are consistent across variables and printing or listing data for documentation of error-checking and error-tracking.

Registration

All users are encouraged to registrate by using the form on www.epidata.dk . By registration you will receive information on updates and help us in deciding how to proceed development - and to persuade others to add funding for the development.

Limits and specification

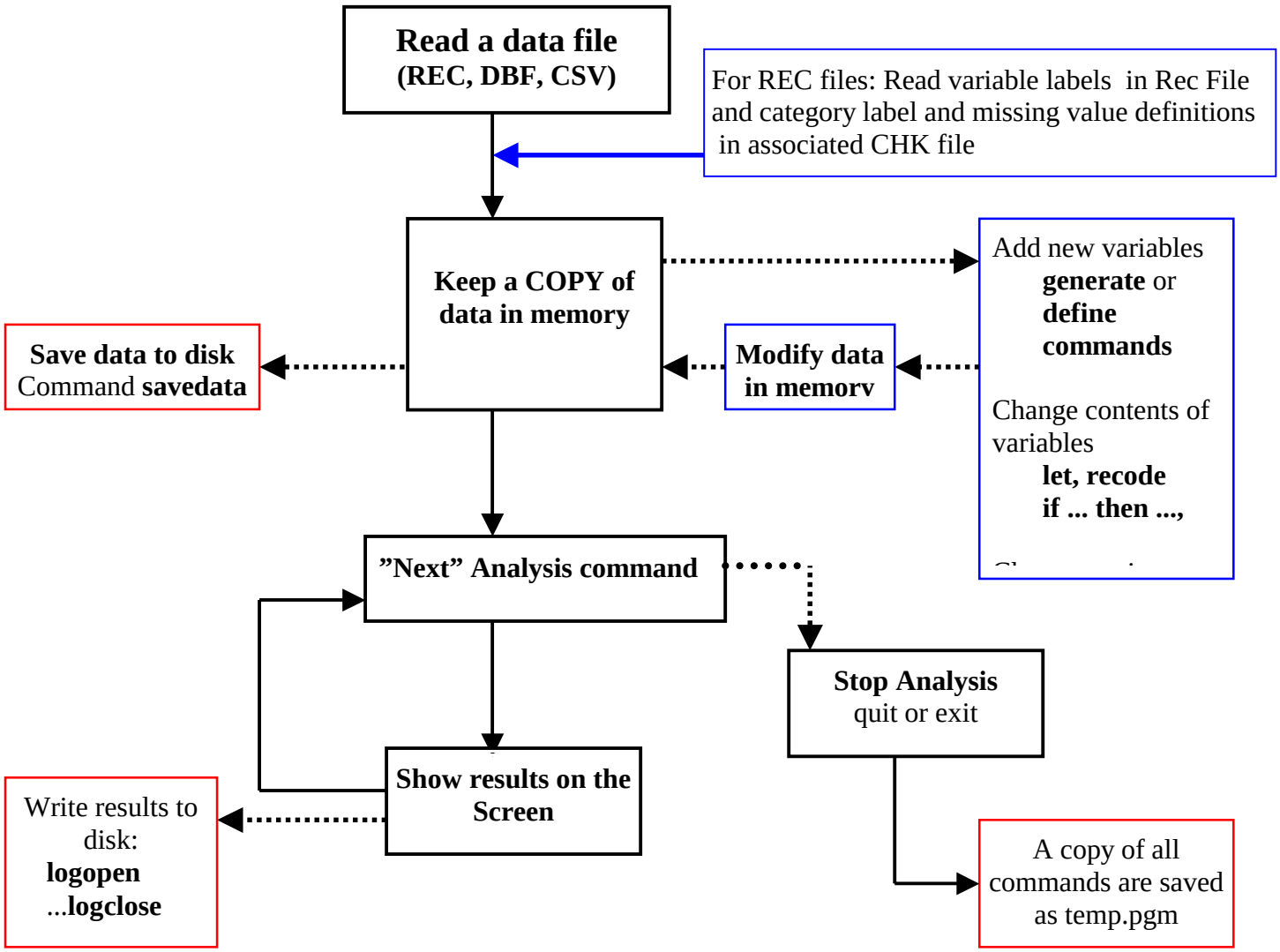
No limit on number of observations. (tested with >100.000). Specification of the data file structure must fit within 999 lines of text. EpiData comes as a few files and only a few extra files are added to accomplish saving of graph files. Options are saved in an ini file. EpiData Analysis is freeware.

Works on Windows 95/98/NT/Professional/2000/XP and Machintosh with RealPc emulator. Tested on more than ten Linux variants based on WINE.

Downloading and installing Download from www.epidata.dk and follow the instructions when you run the installation file. EpiData Analysis will not interfere with the setup of your computer. EpiData consists of one program file and a number of help files in html or pdf format. The program as such cannot be sold for money or service value. It is absolutely free. All translated versions must ALSO be supplied as free. There can be NO charge taken by a web site for downloading of EpiData.

Flowsheet of working with EpiData Analysis

A simplified flowsheet of how EpiData Analysis is working. 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.



Output formatting

EpiData Analysis output is based on the html (hyper text meta language) format. HTML is a structured formatting principle used on all internet pages. EpiData Analysis will comply with the W3C standards. W3C is an international organisation informing about developments and standards on Http://www.w3c.org. For most users knowledge of this is not necessary, but you should know that experienced users can adapt the output as they desire in a style sheet (CSS) containing definitions for size, colour, margins etc. A full documentation of the output principles is available[1] The documentation is installed with the analysis programme.

[1] Lauritsen JM. Output layout definition in EpiData Analysis v1.1. EpiData Association, Odense Denmark, 2005. Http:\\www.epidata.dk\analysisinfo\html output definition.htm

0. Install EpiData Analysis

Get the latest version from Http://www.epidata.dk and install in the language of your preference. The installation and retrieval is fast[2] since the whole size of the programme is small (2.5Mb in total). Simply run the installation file and follow the instructions[3].

1. First run and Setup screen of EpiData:

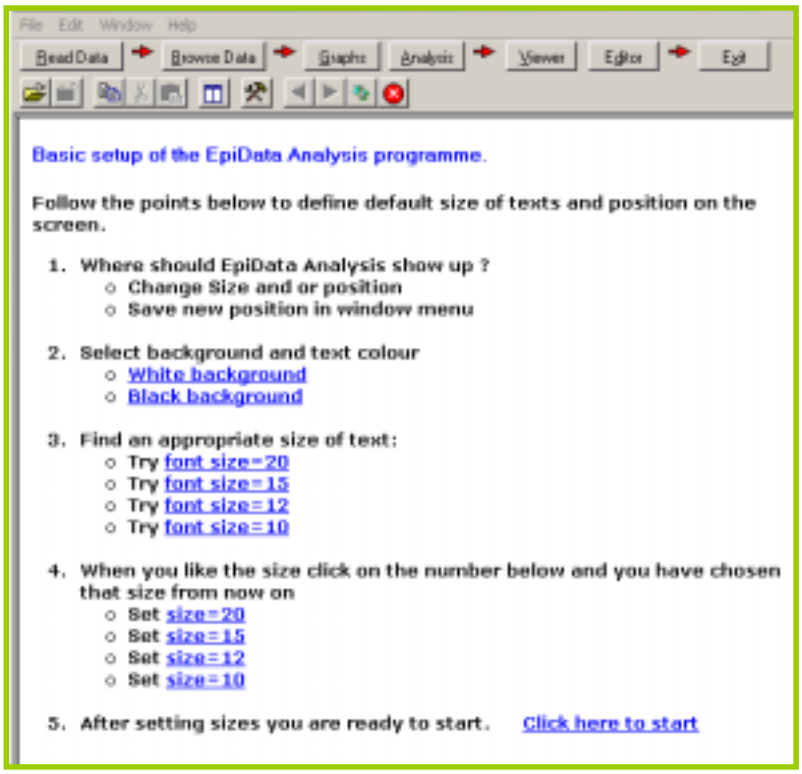
Current version is always shown at the top.

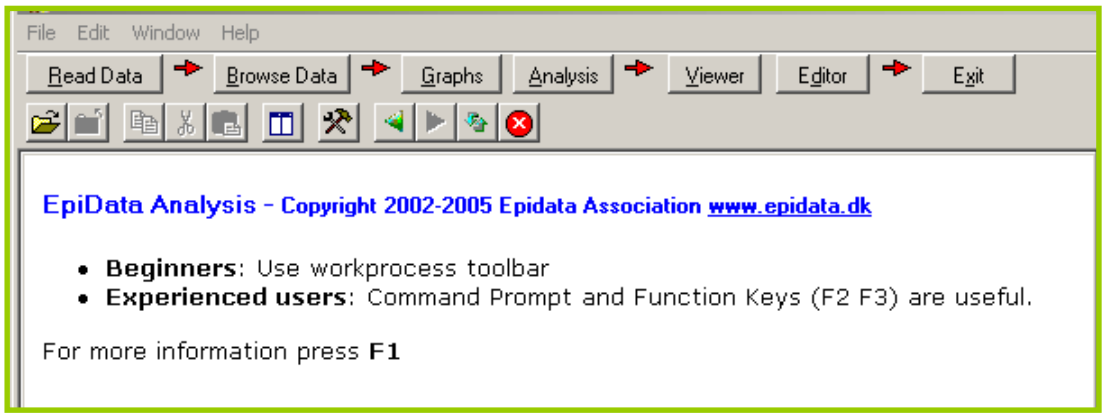
The EpiData Analysis screen has a "standard" windows layout with one menu line, a toolbar, an output window, a command prompt where you write commands, an output window, a statusbar and depending on user action also a data browser, an editor and smaller windows showing available commands, variables in current file and a history of previous commands.

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

B. First time you run EpiData Analysis the opening screen will be a setup screen where you decide overall size of programme, size for fonts you prefer etc. Just follow the numbers:

C. After clicking start, the screen changes.





[2] If you are on a slow modem line you might not agree to "fast", but in comparison to many programmes this is a small size

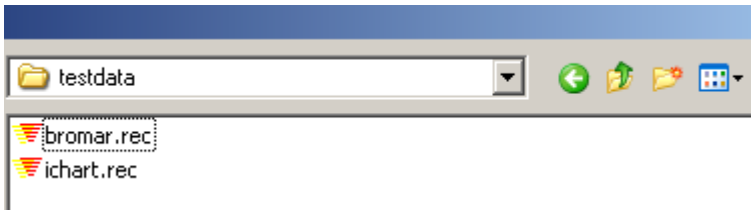
[3] If you are not familiar with installing programmes get help from someone experienced. In some networks you might have to ask a supervisor first. However you can always install on a "temp" folder for yourself as a user.

2. First trial with programme:

To give you an idea of how the programme works we will now try the "Process Toolbar":

Press the "Read Data" button to the left on your screen and point your disk selector to where you installed the EpiData Analysis programme in the subfolder "testdata".

Select "bromar.rec"
(possibly shown only as "bromar").
by double clicking and notice the information in the output window, How many records and fields ?



Data

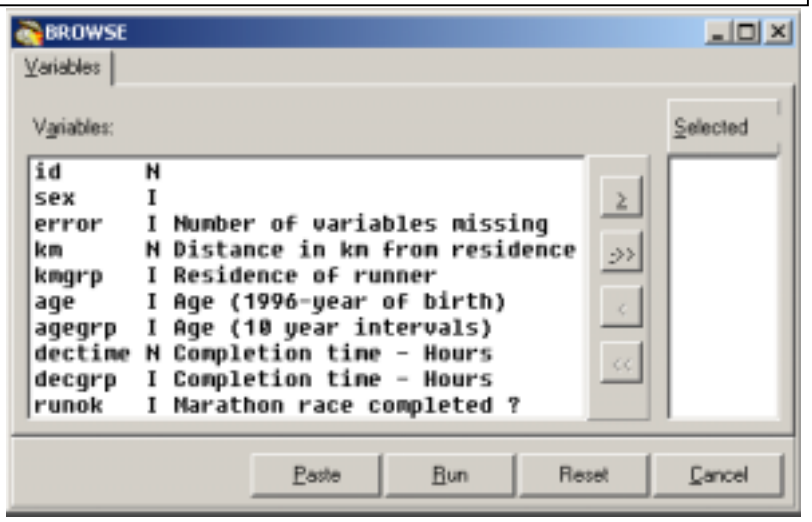
A copy of data is red into memory for a Marathon race.

```
. read

Loading data C:\Programmer\EpiDatab36\testdata\bromar.rec, please wait..

File name : C:\Programmer\EpiDatab36\testdata\bromar.rec
Marathon data - 1995 across bridges from Funen and
Fields: 10 Total records: 4027 Valid records: 4027
```

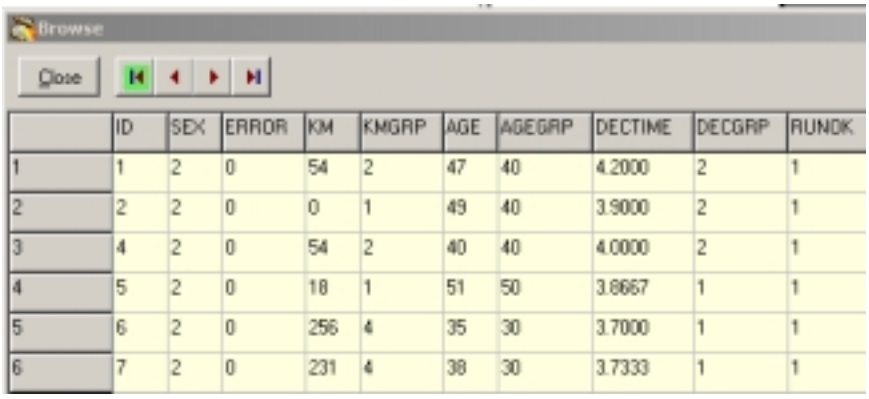
Browse data

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

So on the process tool bar press the "Browse" button, and you get this dialog box:
Select any number of variables or all by the ->> and choose "Run".

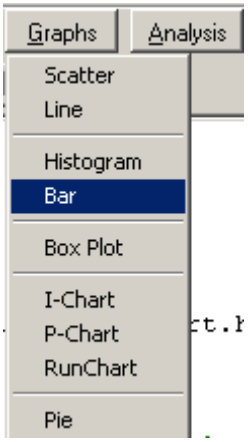


A data browser window is shown. Look at the data. Resize the window to cover right half of the screen and close it again.



	ID	SEX	ERROR	KM	KMGRP	AGE	AGEGRP	DECTIME	DECGRP	RUNOK
1	1	2	0	54	2	47	40	4.2000	2	1
2	2	2	0	0	1	49	40	3.9000	2	1
3	4	2	0	54	2	40	40	4.0000	2	1
4	5	2	0	18	1	51	50	3.8667	1	1
5	6	2	0	256	4	35	30	3.7000	1	1
6	7	2	0	231	4	38	30	3.7333	1	1

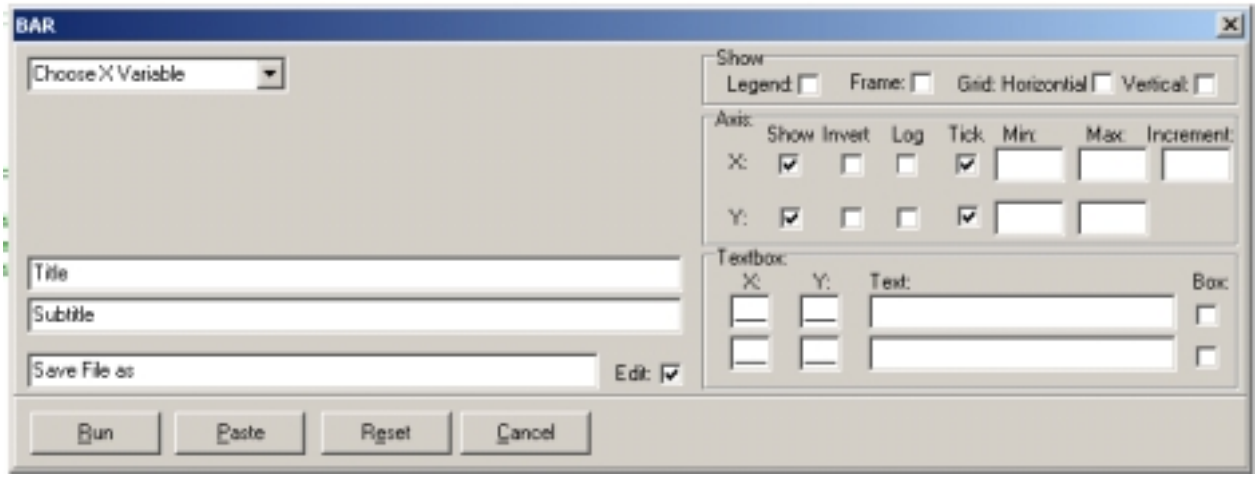
Showing a graph



Now we will try to see some results. Press the **"Graph"** button and you see the small menu, where you choose **"Bar"**

A dialog is shown to specify the graph further:

At this point only worry about the "**Choose X Variable"**. Select **Decgrp** and press **"Run"**

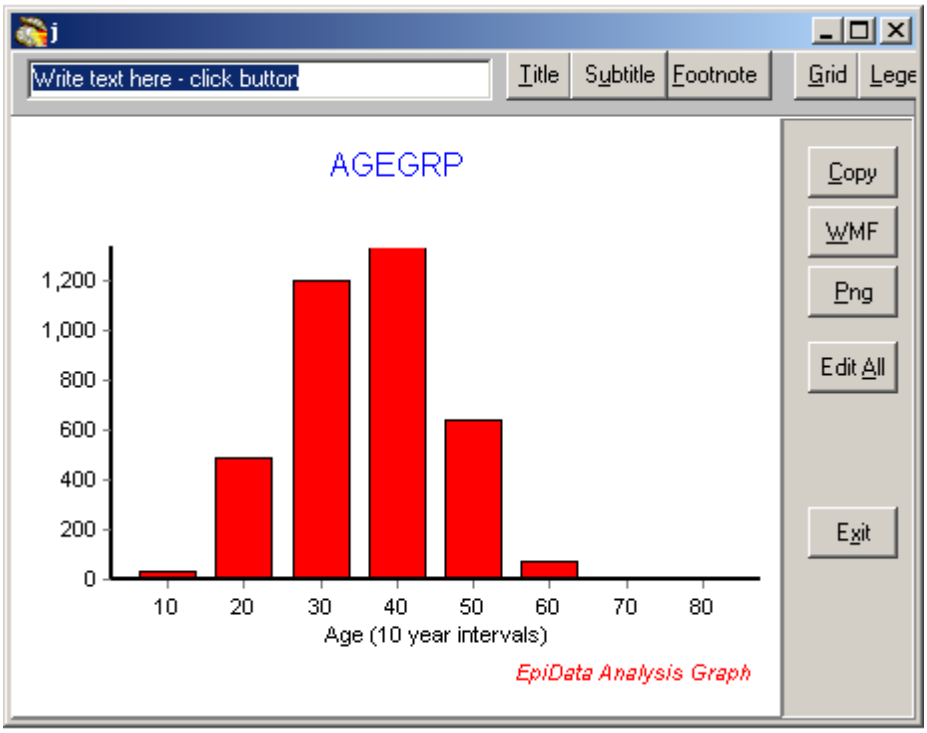


The graph is shown on a special form with some buttons.

Just press "exit"

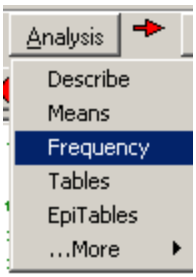
The graph will be added to the output window.

Try with kmgrp or decgrp variables and notice how the labels are shown at the buttom of the graph..



Showing a table of frequencies

On the ”process toolbar” try now analysis:

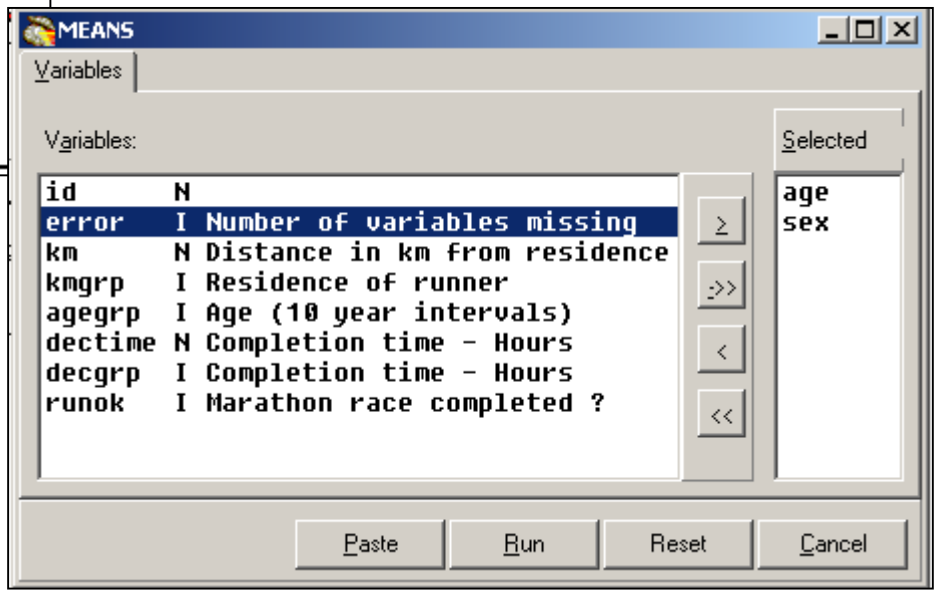


Select sex and agegrp variables in the dialog which is shown and press ”run”, after which you will see these tables:

```
. FREQ agegrp  sex
```

AGEGRP	No.	%
10	35	0.92
20	489	12.92
30	1198	31.64
40	1337	35.31
50	642	16.96
60	75	1.98
70	9	0.24
80	1	0.03
Total	3786	100%

SEX	No.	%
Female	490	12.17
Male	3537	87.83
Total	4027	100%



Showing means of age by time

Choose in Analysis instead ”Means” and ”run”

Are males or females on average the youngest ?

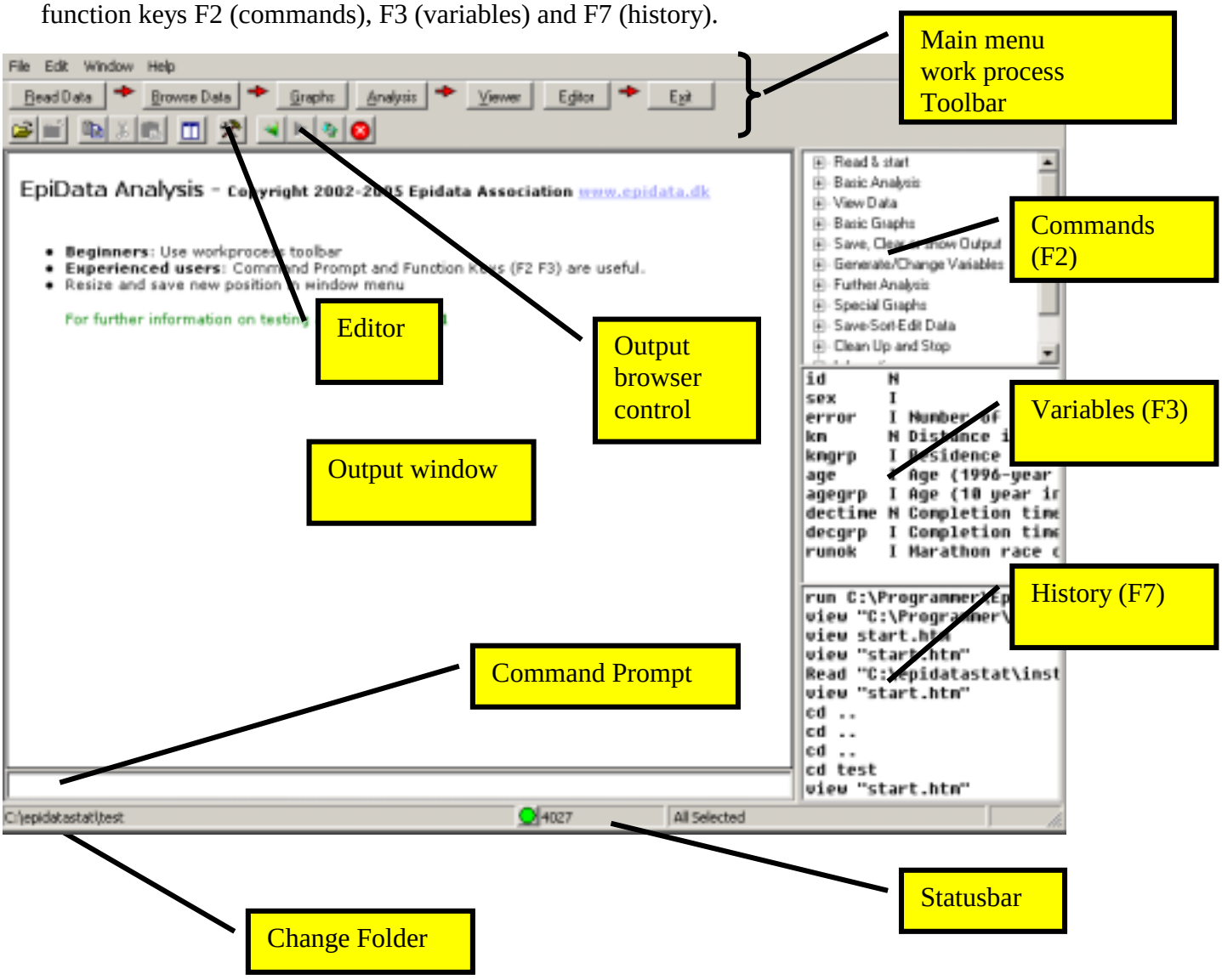
Age (1996-year of birth)

SEX	Obs.	Sum	Mean	Variance	Std Dev	(95% CI	mean)	Std Err	
Female	463	20094.0	43.40	77.97	8.83	42.59	44.21	0.41	
Male	3323	133646.0	40.22	97.60	9.88	39.88	40.55	0.17	
SEX	Minimum	p5	p10	p25	Median	p75	p90	p95	Max
Female	19.00	28.00	30.00	38.00	44.00	50.00	53.00	56.00	70.00
Male	16.00	24.00	27.00	33.00	40.00	47.00	53.00	56.00	84.00

3. Which elements are on the screen ?

You have now acquainted your self with the programme and working with the **process toolbar. On this page you see** a complete picture.

To see a picture like this: Start EpiData Analysis, read a datafile into memory and press function keys F2 (commands), F3 (variables) and F7 (history).



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

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. Basic aspects of epidata follows the Epi Info version 6 manuals. This is available from the Epi Info site: http://www.cdc.gov/Epi Info/
3. 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 improval please discuss with the EpiData-list available at http://lists.umanitoba.ca/mailman/listinfo/epidata-list

Suggested citation of EpiData Analysis program:

General reference for Analysis and EpiData DataEntry is:
Lauritsen JM. (Editor) EpiData Data Entry, Data Management and basic Statistical Analysis System. Odense Denmark, EpiData Association, 2000-2005. (Available from Http://www.epidata.dk).

EpiData Analysis is based on a number of individual contributions:

JM.Lauritsen (2001-) Coordinator. Design, interface, implementation, documentation, testdesign, programming.
Mahmud S. Design and Implementation of core parsing principles and modules for EpiData Analysis. 2003 (source code).
M.Bruus(2004-). Programming and specification. Programming: T.Christiansen(2004-)
Lauritsen JM. Output design definition in EpiData Analysis. EpiData Association, Odense Denmark, 2005. (Contributor J. Hockin) Http:\\www.epidata.dk\analysisinfo\html output definition.htm
Development and Specification of SPC modules in collaboration with Gruk, Norway: V.Høgli (2004), B. Nyen (2004-)

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.