

Getting Started...

Epi Info

Version 6

Mark Myatt



First published in 1991 by Brixton Books
Unit K, Station Building
Llanidloes, Powys SY18 6EB, UK

This edition published 1997

Copyright © 1991, 1997 by Mark Myatt

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without the prior permission of the publisher, nor be otherwise circulated in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

A CIP catalogue record for this book is available from the British Library

ISBN: 1873937016

Trademarks mentioned in this book belong to the following organisations:

Softkey Software Products: WordStar

Microsoft: MSDOS

CONTENTS

Getting Started	1
EPED and Word Processing	
EPED and Word Processing	2
Entering Text	4
Saving and Printing Text	6
Correcting Text	7
Searching for Text	9
Automatically Replacing Text	10
Moving, Copying and Deleting Text	11
Fancy Printing	13
Data Files	15
Survey Processing in EpiInfo	16
Cases, Variables, Questionnaires	17
Creating a Questionnaire File	
Creating a Questionnaire File	19
Variable Names	21
Variable Types	22
Entering and Editing Data	
Entering Data	23
Editing Data	24
Checking Data	25
GIGO	25
The CHECK Program	26
Setting Up Checks	27
Data Analysis	
The ANALYSIS Program	28
Entering Commands	29
Choosing a Data File	30
Examining Data	31
Tables and Statistics	32
Stratified Analysis	33
Continuous Variables	34
Working with Variables	35
Producing Graphs	36
Programming ANALYSIS	37
Oswego County Revisited	40

GETTING STARTED

Information

EpiInfo is a multi-purpose computer program designed for epidemiological researchers and public health personnel. EpiInfo consists of a series of computer programs designed to help with survey design, simple statistics and sample size calculations, data entry, data checking, data analysis, questionnaire design, and report writing.

To use this book you should be seated at your computer and follow the step by step instructions to complete a task.

On different computer systems there will be different ways of starting EpiInfo. On most, however, it will be sufficient to issue the command **epi** or **epi6** at the MSDOS prompt and press the key.







Some of the exercises in this book use the data file **OSWEGO.REC** which is supplied with EpiInfo. If you cannot find this file it is likely to be stored in the **\epi6** directory on the hard disk of your computer.

Instructions

- Start EpiInfo. On different computer systems there will be different ways of starting EpiInfo. On most, however, it will be sufficient to issue the command **epi** or **epi6** at the MSDOS prompt and press the key.


EPED

Instructions

- Start EpiInfo.
- Select **Programs** from the EpiInfo top-level menu. To select an item from the top-level menu you must first move the highlighted bar using the  and  keys. When the highlighted bar is over the menu item you want to select (in this case **Programs**) press the  key to select it. A drop-down menu will then appear.
- Select **EPED** from the drop-down menu. To select an item from a drop-down menu you must first move the highlighted bar using the  and  keys. When the highlighted bar is over the name of the program you wish to use (in this case **EPED**) press the  key to select it. The selected program will then start.

Information

Once **EPED** is started the screen will be blank except for the top and bottom lines of the screen.

The top line of the screen is called the *menu bar*. All the commands available within **EPED** are accessed from the menu bar. To select an item from the menu bar press the associated function key. For example, to access commands that deal with files (e.g. loading and saving files) you would need to press the  key. All commands are selected from menus of the ‘highlighted bar’ type.


The bottom line of the screen is called the status line and conveys information about the document being edited. A typical status line might look like this:

1 UNTITLED EPED 331184 P1 L1 C1 WW Ins

UNTTLED is the name of the file being edited (it is ‘untitled’ because it has not yet been saved to disk and does not have a *filename*).

The large number (**331184**) is the amount of free memory available to EPED.

The numbers preceded by the letters **P**, **L**, and **C** are the current page, line, and column.

Wordwrap (**WW**) means that text will automatically be ‘wrapped’ onto the next line when it reaches the edge of the screen. This means that you do not need to press  at the end of each line.

Insert (**Ins**) means that what you type will not type over what is already on the screen but will be inserted into the text.

Also on the screen you will see a small flashing line or block. This is called the *cursor*. Whatever you type at the keyboard will appear on the screen at the cursor position.

SETTING UP EPED AS A WORDPROCESSOR

Instructions

- Select the **Set** menu by pressing **F6**.
- Point to the **WW/TXT/QES Mode** option using the arrow keys.
- Press the space bar until the mode reads **WW**.
- Press the **ESC** key to close the menu and return to the editing screen.

Information

EPED performs three separate functions:

1. Wordprocessing and Report Writing
2. Creating questionnaires (data entry screens) for use with the **ENTER** program
3. Editing command files for use with the **ANALYSIS** program

Each of these functions require different text editing features. This means that before you start entering text you have to tell **EPED** how to behave.

For wordprocessing work you should set **WW/TXT/QES Mode** to **WW**.

For creating data entry screens you should set **WW/TXT/QES Mode** to **QES**.

For **ANALYSIS** command files and general editing of text files you should set **WW/TXT/QES Mode** to **TXT**.

The current **EPED** mode is shown on the status line.

The choice of whether to use **EPED** in *Insert* or *Overtyp*e mode is up to you. If you are new to wordprocessing it is best to stick to *Insert* mode as this makes it difficult to lose text by typing over it.

Helpful Tips

You can move between *Insert* and *Overtyp*e modes by using the **Set F6** menu or by pressing the **INS** key.

If you need help on a particular **EPED** command then first select the menu on which the command is located, point to the particular command and press **F1**. **EPED** will respond with help for that command. Press **ESC** to start using **EPED** again.

EPED will respond to WordStar commands. Some WordStar dot commands are supported but the dot is replaced with the @ character (e.g. you can type @PA on a blank line to start a new page).

ENTERING TEXT

Instructions

- On the blank screen copy the text from the example text on the next page. Type in the usual way, holding down **[SHIFT]** for a capital letter.
- Remember not to press **[ENTER]** at the end of each line. Only press **[ENTER]** at the end of each paragraph or to add blank lines.
- If you make a mistake when typing move back to the mistake with the arrow keys and press **[DEL]**. The offending character will be removed. Type the correct letter.
- Don't place the title in the centre of the screen using **[SPACE]** or **[F4]**. Wait until you have typed the example text. Use the arrow keys to move to the title line and press **[F4]** to select the **Text** menu. From this menu select the option **Centre current line**. This will place the title in the centre of the line.
- If after correcting text the layout of the text looks a bit odd then move to the beginning of the paragraph using the arrow keys and select the **Text** menu by pressing **[F4]**. Select the option **Format paragraph**. The text will be reformatted.

Helpful Tips

Press the keys lightly. If you hold a key down for too long it will be typed over and over again (repeated).

You can also use the back-space key to correct mistakes. The **[DEL]** key removes the character at the cursor position. The back-space key removes the character to the left of the cursor position.

If you select the wrong command or menu you can cancel it by pressing the **[ESC]** key.

It is best to leave formatting information until you have finished typing the document. This way you can make decisions about the way a document should look in the context of the whole document. It will also help you be consistent in the design and layout of your documents.

EXAMPLE TEXT

Methods of Collecting Data

The methods of collecting information may be broadly classified as follows:

OBSERVATION. The use of techniques varying from simple visual observation to those requiring specialist skills (e.g. clinical examination), sophisticated equipment or facilities such as radiological, biochemical, or microbiological examinations.

INTERVIEWS AND SELF-ADMINISTERED QUESTIONNAIRES. Interviews may be more or less structured. In weakly structured interviews the investigator follows leads as they arise (as in history taking in clinical medicine). In a highly structured interview, questions are decided in advance and are asked in the same order for each respondent. With a self-administered questionnaire, the respondent reads the questions and answers them himself. The highly structured method is preferred since data will be in a standard form accessible to statistical analysis. Less structured interviews may be useful in preliminary surveys.

DOCUMENTARY SOURCES. Clinical records, birth and death certificates, published mortality statistics, census tables, &c. Information gathered from such sources is often called secondary data.

SAVING AND PRINTING TEXT

Instructions

- Select **Save** by pressing **F9**. When prompted give the filename **STARTING.TXT** and press the **ENTER** key. The document will be saved to disk. The first time you save a file you may be asked **Save formatting info?...Y/N**. If you press **Y** the current mode, margins, page length, &c. will be saved with the file and be restored when the file is loaded.
- Check that the printer is ON, that the 'On-line' or 'Ready' light is ON, and that paper is in the printer.
- Select the **Print** menu by pressing **F5** and select the **Print file now** option. The document should print.
- Select **Done** by pressing **F10**. This will return you to the EpiInfo top-level menu.

Helpful Tips

The document has been saved to disk. It can be retrieved using the filename that you specified. When naming files choose a sensible and easily recognizable name. The conventions for naming files are the same as for MSDOS (i.e. 8 characters for the name and 3 characters for the extension separated by a full stop).

Try to be consistent in your use of extensions. Try to use the extensions .TXT or .DOC for wordprocessed files, .QES for questionnaire files (this extension is added automatically when **EPED** is in QES mode), .CHK for **CHECK** files (this extension is added automatically by **CHECK**), and .PGM for **ANALYSIS** command files.

Filenames must be unique. If you specify the name of a file that already exists then **EPED** will ask if you want to *overwrite* the file. If you are not sure that you want to overwrite the file then specify a different filename.

When working on a long document it is best to save your work frequently - especially if where you work is prone to fluctuations or cuts in the electricity supply. You can save a document by pressing **F9**. If you do not save your document and the power fails, you will lose your work.





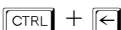



Before issuing the print command check that the printer is connected to your computer and is 'On-line' or 'Ready'. If you try to print a document with a printer not connected or switched off then **EPED** may 'crash' and you may lose your work.

EPED requires that you save your document before printing. If you have not saved the document **EPED** will prompt you to save the file before printing.

CORRECTING TEXT

Instructions

- Start **EPED** by selecting it from the **Programs** menu. Once **EPED** starts select the **File** menu by pressing **[F2]**. Select the **Open file this window** option. Type the filename **STARTING.TXT** and press **[ENTER]**. The text you typed appears on the screen.
- Using the keys shown below, move the cursor around the text, adding and changing the words shown in the example on the next page:.

	One character left
	One character right
	One line up
	One line down
	Beginning of previous word
	Beginning of next word
	Beginning of line
	End of line

- Check your work carefully and correct any mistakes you find.
- Save the file by pressing **[F9]**.

Helpful Tips

In addition to these keys you can use the **[PG UP]** and **[PG DN]** keys to move to the previous and next page of a document. **[CTRL] + [PG UP]** and **[CTRL] + [PG DN]** move to the top and bottom of the document.

To move to the top or bottom of the screen use **[CTRL] + [HOME]** and **[CTRL] + [END]**.

You can only move the cursor over text and lines that have already been typed and not over empty space at the end of the document.

If you delete something by mistake then choose the **Undo this line** option from the **Text [F4]** menu. This will restore the characters you deleted.

If after correcting text the layout of the text looks a bit odd then move to the beginning of the paragraph using the arrow keys and select the **Text** menu by pressing **[F4]**. Select the option **Format paragraph**. The text will be reformatted.

If you cannot remember the name of the file you want to retrieve then press **[ENTER]** when prompted for a filename. **EPED** will respond with a menu listing files in the current directory.

EXAMPLE TEXT

Methods of Collecting Data

The methods of collecting **data** may be broadly classified as follows:

OBSERVATION. The use of techniques varying from simple visual observation to those **needing special** skills (e.g. clinical examination), sophisticated equipment or facilities such as radiological, biochemical, or microbiological examinations.

INTERVIEWS AND SELF-ADMINISTERED QUESTIONNAIRES. Interviews may be more or less structured. In weakly structured interviews the **interviewer** follows leads as they arise (as in history taking in clinical medicine). In a highly structured interview, questions are decided in advance and are asked in the same order for each respondent. With a self-administered questionnaire, the respondent reads the questions and answers them himself. The highly structured method is preferred since data will be in a standard form accessible to **computer based storage and** statistical analysis. Less structured interviews may be useful in preliminary surveys.

DOCUMENTARY SOURCES. Clinical records, birth and death certificates, published mortality statistics, census tables, &c. Information gathered from such sources is often **referred to as** secondary data.

Changes to the original example text are shown in ***bold italics***.

SEARCHING FOR TEXT

Instructions

- Move the cursor to the beginning of the document.
- Select the **Find** menu by pressing **[F7]**.
- Select the **Find..** option.
- Type **clinical** and press **[ENTER]**.
- Type **w** and press **[ENTER]**. **EPED** will search forward through the document until it finds the text 'clinical' and highlight it.

Helpful Tips

You can search for the next occurrence of a word by selecting the **Repeat Last Find/Replace** option from the **Find [F7]** menu.

In the above exercise you typed **w** when asked for options. The **w** signified that you wanted to search for a single word and not part of a longer word. The options available are:

- w** - whole words
- c** - the search will be case sensitive
- g** - search the entire file starting at the beginning (**g** = global)
- u** - search backward (upward) from cursor
- b** - limits search to marked block

Options can be used in any combination. The options **wc** instructs **EPED** to search for whole words taking into account the case of letters.

Searches are always forward from the cursor unless the **u** (upwards) or **g** (global) option is specified.

AUTOMATICALLY REPLACING TEXT

Instructions

- Move the cursor to the beginning of the document.
- Select the **Find** menu by pressing **F7**.
- Select the **Find and Replace..** option.
- Type **sophisticated** and press **ENTER**.
- Type **complicated** and press **ENTER**.
- Type **w** and press **ENTER**. **EPED** will search forward through the document until it finds the text 'sophisticated' and highlight it. Type **Y** to the **Replace? (Y/N/A/Q) :** prompt. **EPED** will replace the word.

Helpful Tips

You can repeat the **Find..** or **Find and Replace..** operation by selecting the **Repeat Last Find/Replace** option from the **Find F7** menu.

In the above exercise you typed **w** when asked for options. The **w** signified that you wanted to search for a single word not a part of a word. The options available are:

- w** - whole words
- c** - the search will be case sensitive
- g** - search the entire file starting at the beginning (**g** = global)
- u** - search backwards (upwards) from the cursor
- n** - replace without asking
- b** - limits search to marked block

Options can be used in any combination. For example the option **wng** instructs **EPED** to search the whole file (**g**) for whole words (**w**) ignoring the case of letters and replace them all without asking (**n**).

Searches are always forward from the cursor unless the **u** (upwards) or **g** (global) option is specified.

The **Replace? (Y/N/A/Q) :** prompt also alters the operation of the **Find and Replace..** function:

- Y** - Replace highlighted word
- N** - Do not replace highlighted word
- A** - Replace all occurrences of the word
- Q** - Quit **Find and Replace..**

Instructions

- Practice using the **Find..** and **Find and Replace..** options from the **Find F7** menu. Try experimenting with the various options.

MOVING, COPYING AND DELETING TEXT

Information

EPED, in common with most wordprocessors, uses the concept of *blocks* when moving, copying and deleting large chunks of text. Working with blocks always involves issuing three separate commands. These are:

1. Mark the start of the block
2. Mark the end of the block
3. Perform operation (e.g. *move*, *copy* or *delete* the marked block)

The block commands are all found on the **Blk** menu which is accessed by pressing the **F8** key.

Instructions

- Move the cursor to the beginning of the paragraph starting **INTERVIEWS AND . . .**
- Select the **Blk** menu by pressing **F8**.
- Select **Begin Block**.
- Move the cursor to the blank line after the paragraph.
- Select the **Blk** menu by pressing **F8**.
- Select **End Block**. The block is highlighted.
- Move the cursor to the end of the document.
- Select the **Blk** menu by pressing **F8**.
- Select **Move block to here**. The block is moved to the end of the document.

Helpful Tips

The block remains highlighted (i.e. selected) even after it has been moved. To deselect a block choose **Hide block display** from the **Blk F8** menu. When you mark the beginning of a new block **EPED** remembers where the previous end-of-block marker was and will highlight up to it. Do not worry about this. Carry on and mark the end of the block as you would normally.

The sequence for selecting a block is simple: Move to start of block / **F8** / **Begin block** / Move to end of block / **F8** / **End block**.

You can also use a combination of the **CTRL** key and the function keys to perform block operations. To see these key combinations hold down the **CTRL** key.

A block may be of any size from a single character to the entire document.

MOVING, COPYING AND DELETING TEXT

Information

The option **Write block to file...** on the **Blk** **F8** menu writes the currently marked block to a disk file. You will be prompted for a file name. If the file already exists you can choose to append to (i.e. add text to the end of the existing file) or overwrite the file. The append feature is useful when editing output from the **ANALYSIS** module. You can select the tables and statistics you intend to include in a report by marking each table as a block and then appending the block to a file. This file can then be used as the basis of a report. When using this command to append blocks to a file make sure that the file you specify is a text file.

The option **Read file in as block** on the **Blk** **F8** menu reads the contents of an entire file into the current document at the cursor position. You will be prompted for the name of the file. Make sure that the file you specify is a text file. The block will remain selected so it can be formatted with the **Format current block** option.

Instructions

- Practice using the block commands to move, copy and delete blocks of text.

Helpful Tips

The commands for moving, copying and deleting blocks of text are accessed from the **Blk** **F8** menu or using the **CTRL** and function key combinations.

The block remains highlighted (i.e. selected) even after it has been moved. To deselect a block choose **Hide block display** from the **Blk** **F8** menu. When you mark the beginning of a new block **EPED** remembers where the previous end-of-block marker was and will highlight up to it. Do not worry about this. Carry on and mark the end of the block as you would normally.

The option **Format current block** on the **Blk** **F8** menu reformats text in a selected block. This is useful because the **Format paragraph** option on the **F4** **Text** menu will only reformat text one paragraph at a time. A quick way to reformat an entire document is to mark the whole document as a block and then use the **Format current block** option. The key sequence for doing this is **CTRL** + **PG UP** / **F8** / **Begin block** / **CTRL** + **PG DN** / **F8** / **End block** / **F8** / **Format current block**.

FANCY PRINTING

Information

EPED lets you print in several type styles depending on the type of printer you use. It is simplest to treat typestyles as a block operation. Select the text you wish to enhance (e.g. with italics, bold, &c.) as a block using the Move to start of block / **[F8]** / **Begin block** / Move to end of block / **[F8]** / **End block** command sequence. Then select the **Text** menu by pressing **[F4]**. From the **Text** menu select **Style of type**. A second menu will appear. Select the desired typestyle.

Instructions

- Experiment with changing the styles of type in your document. Use all of the typestyles listed in the **Style of type** menu.
- Save the document by pressing **[F9]**.
- Print the document so you can see what each style looks like when printed.
- Return to the EpiInfo top-level menu by pressing **[F10]**.

Helpful Tips

It is best to leave formatting information until you have finished typing in the document. This way you can make decisions about the way a document should look in the context of the whole document. It will also help you be consistent in the design and layout of your documents.

EPED places some non-printing characters (called *control codes*) directly in front of and behind text changed to bold, underline, italic &c. These control codes can be deleted by placing the cursor in the appropriate spot and pressing **[DEL]**. When the cursor is placed over a control code it will change from a flashing line to a block. The control character (e.g. **^B** for bold) will be shown on the far right of the status line on the bottom of the screen. It is important that you delete **both** sets of type style information in this way. Control characters appear in the text but do not print. Each style of type has different control characters:

Style of Type	Screen	Status Bar
Bold	Block	^B
Double	Shaded block	^D
Italic	Slash (/)	^N
Underline	Underscore	^S
Compressed	Small a	^A
Superscript	Corner graphic	^T
Subscript	Corner graphic	^V

Some type styles *may* not print on your printer.

GOOD LUCK

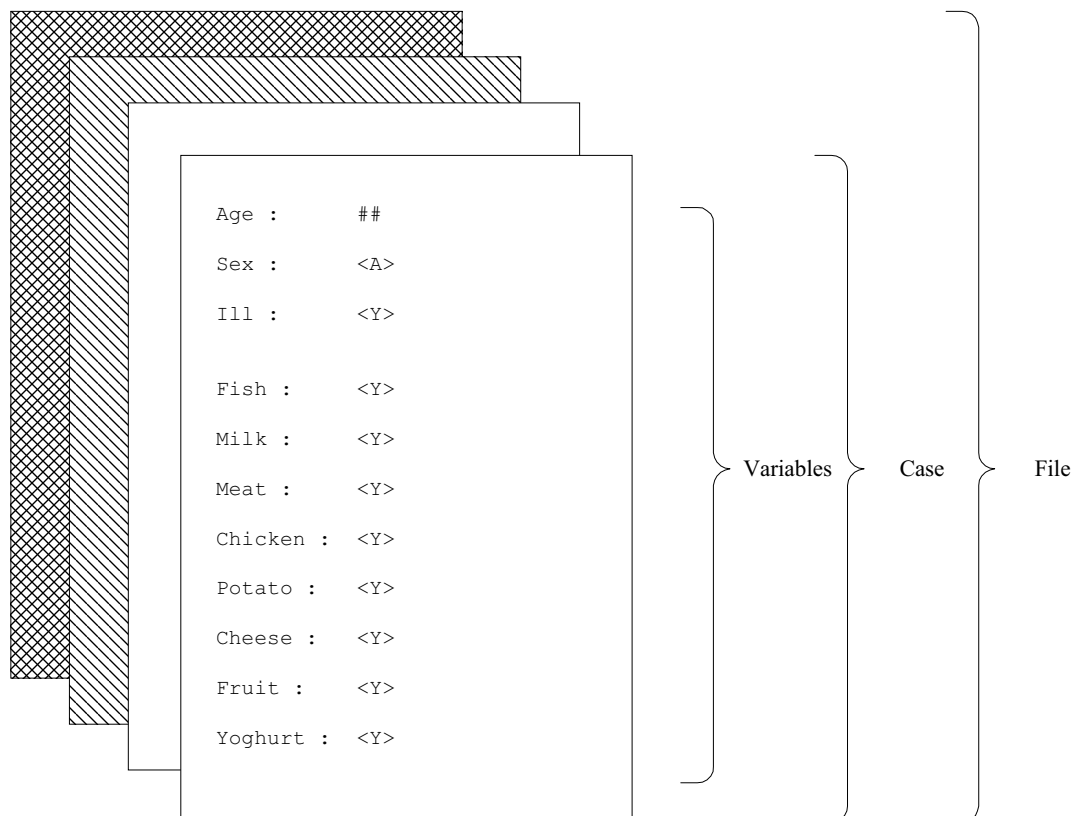
You have used most of the simple editing commands in **EPED**. There are some more advanced features which are discussed in the manual. The way to learn to use a wordprocessor is simply to practice and experiment. You might like to run through the tutorial again to boost your confidence. We will now move on to using **EPED** to create questionnaire files and data entry screens.

DATA FILES

Information

A *data file* is structured in a certain way. Typically a data file will consist of a collection of *cases* (or *records*) which are in turn made up of *variables* (or *fields*) which hold data for each individual case.

Each case is made up of a collection of the same variables but each holding different data:



With EpiInfo a data file is created in two separate steps.

First a *questionnaire* file is created (using **EPED**) which defines the structure of each case (i.e. the name, type, and length of each variable). This file acts as a template for the structure of each record in the data file. The questionnaire file **must** end with the extension .QES.

The **ENTER** program is then used to create a data file which has the extension .REC.

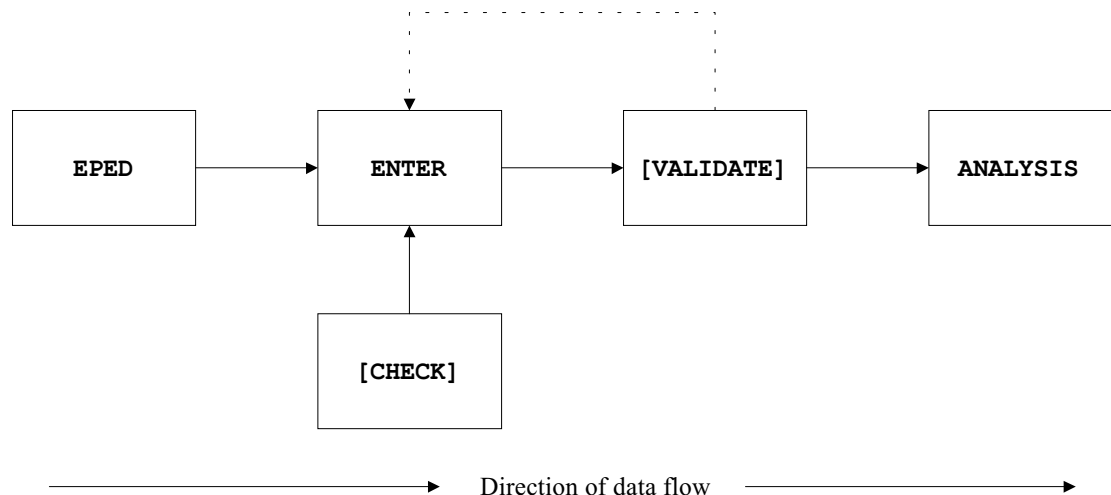
The **ENTER** program is also used to enter data into cases.

Files created in this way can then be analyzed using the **ANALYSIS** program.

SURVEY PROCESSING

Information

With EpiInfo, survey processing is split into several processes each of which are handled by different EpiInfo modules:



The functions of each module are:

EPED is used to create questionnaire files which define the structure of the data and the layout of the data entry screens.

ENTER is used for data entry and editing. The questionnaire file created with **EPED** is used by **ENTER** to create a data file.

ANALYSIS is used to analyze data and produce listings, tables, statistics, and graphs.

In addition to these three main modules there are two other modules that provide data checking functions:

CHECK is used for *interactive* data checking so that data may be checked as it is entered.

VALIDATE provides for the *batch* validation of data files. Data is entered twice (*double-entered*). The program compares the contents of the two sets of data and reports any differences between them. Data can then be edited using the **ENTER** program.

The use of the **CHECK** and **VALIDATE** programs is optional but recommended as they provide a means of ensuring data of reasonable quality.

CASES AND VARIABLES

Information

One of the main uses for the **EPED** text editor is the creation of questionnaire files (data entry forms) for use with the **ENTER** program. Before you can begin to enter data you must first create a questionnaire file using **EPED**.

The first function of a questionnaire file is to act as the data entry form during data entry operations. A questionnaire file can be printed and used as a form for data collection. Such a questionnaire will **not** usually be suitable for data collection. It is better to first create a questionnaire for data collection (data collection form) and edit a copy of the file for use as a data entry form or set of data entry screens. Complex coding information, whilst needed on paper forms, would tend to clutter up data entry screens.

The second function of the questionnaire file is the definition of variables collected for each case. A variable has three attributes:

Attribute	Purpose
Name	Each variable has a <i>name</i> . In the ANALYSIS program statistical analysis is performed by referring to variables by their names. Variable names can be up to ten characters long.
Type	Each variable must be of a certain <i>type</i> . The type you choose to assign to a variable will depend on the type of data you wish it to contain and the type of statistical analysis you would like to perform.
Length	The <i>length</i> of a variable defines how much information a variable may hold. A text variable of length ten will be able to hold up to ten letters and numbers. A numeric variable of length three will be able to hold numbers between -99 and 999.

You specify each of these attributes for each of your variables in the questionnaire file.

CASES AND VARIABLES

Information

Each variable must be of a certain *type*. The type you choose to assign to a variable will depend on the type of data you it is to contain and the type of statistical analysis you would like to perform. EpiInfo allows for many different variable types:

TEXT variables are used for holding information consisting of text and / or numbers. Text variables are useful for holding information such as names and addresses. There is a special form of text variable called the **UPPER CASE TEXT** type. This is the same as the TEXT type but can only hold upper case letters as well as numbers. If you enter lower case text into an UPPER CASE TEXT variable it will automatically be converted to upper case text. You cannot perform mathematical operations with text variables.

NUMERIC variables are used for holding numerical information. They can be used to hold continuous or categorical data.

YES/NO or **LOGICAL** variables are used for holding information which can have two possible states such as whether a person has been ill or not. YES/NO variables can hold either the character **Y** or the character **N**.

DATE variables are used to hold dates. Date variables can be used to hold dates in the American (MM/DD/YY or MM/DD) or European (DD/MM/YY or DD/MM) format. The advantage of using DATE variables over TEXT variables to enter and store dates is that the **ENTER** program will only allow you to enter valid dates.

PHONENUM, **LAST UPDATE**, and **AUTO-INCREMENT**. These are of limited use.

Helpful Tips

When creating your own questionnaire files think carefully about the sort of data you want each variable to hold. If you wish to perform mathematical operations with variables they should be of the NUMERIC type.

Some investigators prefer to use numeric variables to hold all data. Categorical variables are given numerical codes. Data files created in this way may be easier to use with other statistical packages.

Many statistical packages do not support date type variables. If you use date type variables then you may need to convert them to a different format (e.g. calendar month in century) before they can be used by other packages. Variables can be changed or *recoded* with the **ANALYSIS** module.

CREATING A QUESTIONNAIRE FILE

Questionnaire files (data entry forms) are created using **EPED**. Variable names are typed onto the screen in the position and order you would like to use when entering data. Next to the variable name are typed special characters that define the type and length of the variable.

The sample questionnaire file is typical of one used to enter data for the analysis of an outbreak of food poisoning. Later you will use this questionnaire file to create a data file and enter data.

Instructions

- Start **EPED**.
- Make sure **EPED** is in QES mode.
- Type the sample questionnaire file shown on the next page.
- Check your work carefully and save the file as **STARTING.QES**.
- Return to the EpiInfo main menu by pressing **F10**.

Helpful Tips

EPED has a feature that makes it easy to define variable types and lengths. On the **Txt** **F4** menu there is an item called **Questions**. If you select this item you will be presented with a menu of variable types. Select the variable type you require. If you choose a type that allows for variables of different lengths (e.g. TEXT, UPPER CASE TEXT, or NUMERIC) you will be prompted for the appropriate information.

When designing your own questionnaire file bear in mind that they must be easy to understand and complete. A questionnaire that looks good on paper will probably look good on the screen. If your study requires a great deal of data entry pay careful attention to the design of the questionnaire file. Complex coding information, whilst needed on paper collection forms, would tend to clutter up the data entry screen.

Each data entry screen is 23 lines long. When **EPED** is in QES mode the page length is set to 23 lines. The start of each screen is marked by dark block characters on each side of the **EPED** screen.

If after creating a data file you find that you need to make some changes to the questionnaire (e.g. layout, adding a new variable, changing the length of an existing variable) you will need to restructure the data file using **ENTER**. To do this start **ENTER** in the usual way and type the name of the existing data file and press **ENTER**. You should then specify the third option (**3. Revise structure of data file using revised .QES**) and press **ENTER**. You will then be prompted for the name of the new (or revised) questionnaire (.QES) file. Type the name of the questionnaire (.QES) file and press **ENTER**. Press **ENTER** again to confirm that you wish to proceed. The **ENTER** program will then restructure your data file to take into account changes in the questionnaire (.QES) file.

SAMPLE QUESTIONNAIRE FILE

Questionnaire for Church Supper - Oswego

Age : ##
Sex : <A>
Time of Supper (24 hour) : ####
Ill? <Y>
{Onset Date} (DD/MM) : <DD/MM>
{Onset Time} (24 hour) : ####

Food History (Mark Y or N):

Baked Ham : <Y>
Spinach : <Y>
Mashed Potatoes : <Y>
Cabbage Salad : <Y>
Jello : <Y>
Rolls : <Y>
Brown Bread : <Y>
Milk : <Y>
Coffee : <Y>
Water : <Y>
Cakes : <Y>
{Vanilla} Ice Cream : <Y>
{Chocolate} Ice Cream : <Y>
Fruit Salad : <Y>

VARIABLE NAMES

Information

EpiInfo will take the first ten characters that you provide for each variable name. The text 'Cabbage Salad' will be used to create a variable called CABBAGESAL.

Punctuation characters are not included in variable names. The text 'Ill?' in the sample questionnaire file will be used to create a variable called ILL.

EpiInfo may also modify some of the variable names you supply. In the sample questionnaire file the text 'Time of Supper' will be used to create a variable called TIMESUPPER.

Helpful Tips

Variable names enclosed in braces {} will be used in preference to those normally generated. In the sample questionnaire {Vanilla} is used to create a variable VANILLA instead of VANILLAICE. The same technique is used for variables CHOCOLATE, ONSETDATE, and ONSETTIME. The brace characters {} take up room in the .QES file but do **not** appear on the data entry screens. This is why the variable definition characters (<DD/MM>, ####, and <Y>) used to define the type and length of these variables are placed two spaces to the right of those used to define the type and length of the other variables in the sample questionnaire (.QES) file.

It is a good idea to create variables with unique names. If you use duplicate variable names EpiInfo will append a number to the end of the variable name. If there were two variables named ADDRESS then EpiInfo would create two variables called ADDRESS and ADDRESS2.

Try to give your variables short, easily remembered, and relevant names. This will make it easier for you to use the **ANALYSIS** program where you have to refer to variables by name. For long questionnaires it is often easier to use question numbers as variable names.

The **CHECK** module also provides a way of changing the names of variables. You can only use this module after you have used **ENTER** to create a data file.

VARIABLE TYPES

Information

NUMERIC variables are defined using the **#** character. A variable defined as **###** can hold 3 digits. If a decimal point is given then the variable will be in *fixed decimal* format. A variable defined as **###.##** can hold numbers between -99.99 and 999.99. Numerical variables defined in the sample questionnaire file are AGE, TIMESUPPER, ONSETTIME.

TEXT variables are defined by using the underline character. A variable defined using three underline characters will be able to hold up to three characters and / or numbers. There are no text variables defined in the sample questionnaire file.

UPPER CASE TEXT variables are defined by enclosing an upper case **A** within less-than and greater-than signs. The number of characters between the **<** and **>** signs defines the length of the variable. A variable defined as **<AAAAA>** will be able to hold up to five upper case letters and / or numbers. SEX is the only upper case text variable defined in the sample questionnaire file.

DATE variables are defined by enclosing the date format between less-than and greater-than signs. A variable defined as **<DD/MM/YY>** will be able to hold a long European date. The variable ONSETDATE is defined to hold a short European date in the sample questionnaire.

YES/NO fields are defined by enclosing an upper case **Y** between less-than and greater-than signs (e.g. **<Y>**). The variables ILL, BAKEDHAM, SPINACH, MASHEDPOTA, CABBAGESAL, JELLO, ROLLS, BROWNBREAD, MILK, COFFEE, WATER, CAKES, VANILLA, CHOCOLATE, and FRUITSALAD are defined as YES/NO variables in the sample questionnaire file.

Helpful Tips

When designing your own questionnaire files think carefully about the sort of data you want each variable to hold. If you wish to perform mathematical operations with variables they should be of the NUMERIC type. Always use NUMERIC variables for continuous data such as ages or heights. With categorical variables it is a good idea to keep codes consistent across variables. This will help reduce errors at all stages of a survey.

When analyzing data you may need to create new variables that summarize the contents of existing (i.e. collected) variables. Do not include these variables on the questionnaire file as they can be created and defined within the **ANALYSIS** program.

When creating your own questionnaire files it is a good idea to include a field that is unique to each record. This makes finding records and paper forms easier should you need to query a data item. It also makes it easier for double-entry and validation as records will not need to be entered in the same order for both files.

Some investigators prefer to use numeric variables to hold all data. Categorical variables (such as SEX and the YES/NO variables in the sample questionnaire) are given numeric codes. Data files created in this way may be easier to use with other statistical packages.

UPPER CASE TEXT variables are preferable to simple TEXT variables. UPPER CASE TEXT variables automatically convert lower case text into upper case. This avoids problems caused by inconsistent capitalization of (e.g.) postcodes or state names.

ENTERING DATA

Instructions

- Select **ENTER** from the **Programs** menu
- In response to the request for the data file type **STARTING** and press **[ENTER]**. As this file does not already exist you should specify option **2 (Create new data file from .QES file)** and press **[ENTER]**. You will then be prompted for the name of the questionnaire (.QES) file.
- In response to the request for a questionnaire file enter **STARTING** and press **[ENTER]**.
- Press **[ENTER]** again to confirm that you wish to proceed. The **ENTER** program will use the questionnaire file (**STARTING.QES**) to create a data file (**STARTING.REC**) and display the questionnaire on the screen, ready to receive data.
- Fill in the blanks in the questionnaire (make up some data) until you have completed the form.
- After you have completed a questionnaire form (i.e. entered all the data for one case) the message **Write data to disk? (Y/N/<Esc>)** will be displayed at the bottom of the screen. Answer **[Y]** to this question. A new (blank) questionnaire will be displayed on the screen. Note that record number **2** is displayed in the lower right hand corner of the screen.
- Enter data for several more cases, writing each case to disk.

Helpful Tips

You can only enter the type of data that is shown on the prompt line at the bottom left of the screen. This may be different for each variable and depends on the variable types you specified in the questionnaire (.QES) file.

If you enter data of the wrong type into a variable (or an invalid date into a DATE type variable) then **ENTER** will beep and refuse to accept the bad data. Re-enter the correct data.

Pressing **[ENTER]** on its own (i.e. entering no data into a variable) will set the variable to 'missing'.

The **[↑]** and **[↓]** keys allow you to move between variables.

The **[HOME]** key will move the cursor to the first variable on the questionnaire form. The **[END]** key will cause the **Write data to disk? (Y/N/<Esc>)** message to be displayed.

If you answer **[N]** to the **Write data to disk? (Y/N/<Esc>)** prompt or press **[ESC]** then **ENTER** will not save the case to the disk but present you with the same case which can then be edited.

Pressing **[F10]** returns you to the top-level menu.

EDITING DATA

Information

Data can also be edited after it has been saved to disk.

The **F7** and **F8** keys provide a means of moving from case to case within a data file. The **F7** key moves one case back and the **F8** key moves one case forward.

The displayed case can be edited and then saved to disk by pressing the **END** key.

To continue entering new cases press the **CTRL** + **N** keys.

Instructions

- Use the **F7** and **F8** keys to move through the cases and review your work.
- Edit one of the cases by pointing to a variable with the cursor and changing its contents.
- Save the edited case by pressing the **END** key and replying with **Y** to the **Write data to disk (Y/N/<Esc>)** prompt.
- Return to data entry mode by pressing **CTRL** + **N**.
- Return to the EpiInfo main menu by pressing **F10**.

Helpful Tips

You can find a particular case (or set of cases) using **CTRL** + **F** to place the **ENTER** program in *find mode*.

In *find mode* you can enter data into a questionnaire form that matches the data held in the case(s) you wish to find. The following keys work in find mode:

F2	Find case by record number
F3	Find case(s) that match data entered after find mode was started
ESC	Stop search
F4	Resume search (find next case)

You can edit cases in find mode.

Press **CTRL** + **N** to enter new records.

The **F5** key in **ENTER** allows you to print a copy of the data entry screen, data for a single case, and data for a series of cases.

You can delete a case by pressing **F6** when the case you want to delete is displayed. Data is not deleted but is *marked as deleted* and excluded from data analysis by the **ANALYSIS** program. **F6** is a toggle. Press it once to delete a case. Press it again and the case is 'undeleted'. A deleted case is marked by an asterisk (*****) next to the record number on the bottom right corner of the screen.

GIGO

Information

GIGO is the most important acronym in computer science. It stands for *Garbage In Garbage Out*. Put simply this means if your data is of poor quality then it is certain to be the case that the results of any analysis will be unreliable. For this reason great emphasis is placed on checking data. In order to check data effectively you should be aware of the types of error that can find its way into the data and how to check for them.

Types of error

Transposition. e.g. 39 becomes 93.

Coding errors.

Routing Errors. Interviewer asks wrong questions in wrong order.

Copying errors. e.g. 6 as 0, 0 as O, 1 as l, &c.

Consistency errors. Answers to questions conflict.

Range errors. Answers to questions lie outside of probable, or possible, values.

Methods of Checking the Data

There are four methods of checking data. All four methods should be applied to data as each method has different strengths and weaknesses.

Manual Checking. Apply this test to a small number of completed questionnaires. It provides an initial assessment of data quality and highlights problems with the questionnaire or data collection. Manual checking is particularly useful in detecting routing errors.

Checking during data entry (Interactive Checking). The **CHECK** program allows immediate detection and correction of problems in the data. Interactive checking is useful in picking up range, copying, questionnaire routing and consistency errors.

Checking after data entry (Batch Checking). It is possible to use the **ANALYSIS** program to check for errors in data once all the data has been entered. Producing frequency tables and scatter plots should highlight possible errors in the data. It is also possible to check that data is consistent and pick up range errors.

Validation (aka Verification). This involves the data being entered twice. The resulting sets of data are then compared with each other to see if they are the same. The data is then checked and edited. Validation is useful in picking up copying errors. The **VALIDATE** program in EpiInfo provides validation. You can use **ENTER** to double enter and validate data interactively (option labelled **4. Reenter and verify records in existing data file**).

If you use **VALIDATE** to compare files then you should ensure that data are entered in the same order in both files or that each case has a unique identifying variable. If you use **ENTER** to reenter and verify files each case must have a unique identifying variable.

CHECK

Information

The **ENTER** program provides a limited form of data checking. It will allow you to enter only numbers into NUMERIC variables, valid dates into DATE variables, and **Y** or **N** responses into YES/NO variables. Although you may use the **ENTER** program on its own it is far better to use it in conjunction with the **CHECK** program. **CHECK** contains some features that allow for interactive error checking, automatic coding of variables, and control over questionnaire routing. These functions are listed below:

Must-enter fields. **CHECK** allows you to specify that certain variables on the questionnaire must be filled with a value other than missing.

Legal values. The input must match one of the specified list of values. The variable can be left blank unless it is designated as a *must-enter* variable.

Range checks. The contents of a variable must lay between two bounding values. TEXT and UPPER CASE TEXT variables may use range checking based on alphabetical order.

Repeat variable. The variable will automatically hold the value entered for the previous case. This is useful for data that seldom changes (e.g. location codes).

Automatic coding. The entry is compared against a pre-entered table of codes and phrases. If a match is found another variable can be set to the code value.

Conditional jumps. **CHECK** allows you to check a variable for certain values that if found will cause the cursor to jump to a specified variable. If the test fails the cursor moves to the next variable.

ENTER and **CHECK** provide one type of data checking called *interactive checking* (i.e. checking the validity of data as it is entered). The **VALIDATE** program allows for *batch* (i.e. all cases at once) checking of data which has been *double-entered*. Checks for consistency of data and cross-case checks can be performed with the **ANALYSIS** program.

SETTING UP CHECKS

Instructions

- Select **CHECK** program from the **Programs** menu.
- Enter the filename **STARTING** and press **[ENTER]**. The **STARTING** data entry screen appears on the screen (as with the **ENTER** program) but with a different set of function key prompts at the bottom of the screen. This is not a data-entry screen. Do not attempt to enter data now.
- Type **1** into the AGE variable. Press **[F1]** to make this the minimum value allowed in this variable.
- Type **90** into the AGE variable. Press **[F2]** to make this the maximum value allowed in this variable.
- Type **M** into the SEX variable. Press **[F6]** to make this a legal-value.
- Type **F** into the SEX variable. Press **[F6]** to make this a legal value.
- Type **U** (Unknown) into the SEX variable. Press **[F6]** to make this a legal value.
- With the cursor still in the SEX variable press **[F4]** to make this a must-enter variable.
- Type **N** into the ILL variable and press **[F7]**. Move the cursor to the BAKEDHAM variable and press **[F7]** again. During data-entry the cursor will jump automatically to the BAKEDHAM variable when **N** is entered into the ILL variable.
- Press **[F10]** and confirm that you wish to save the checks.
- Use the **ENTER** program to enter some more data into **STARTING.REC**. Verify that the range, legal-value, and jumps work correctly.
- Return to the EpiInfo main menu by pressing **[F10]**.

Information

The checks are stored in a file with the extension **.CHK** with the same filename as the original data file. The checks set up above will be stored in a file called **STARTING.CHK**. This is a plain text file containing instructions that are used by the **ENTER** program. This file may be edited using a text editor such as **EPED**.

If you need to use a lot of **CHECK** commands it can be easy to lose track of what you are doing. Pressing **[F9]** displays all the **CHECK** commands you have entered for a particular field.

CHECK can also be used to change the name of a variable. Place the cursor at the variable you want to change. Press **[F9]** and then **[F1]**. You will be prompted for a new variable name. After you have entered the new name press **[ESC]** to return to the main **CHECK** screen.

ANALYSIS

Information

ANALYSIS is the data analysis module of EpiInfo. Using **ANALYSIS** you can use simple commands to produce lists, tables, frequencies, statistics, and graphs. The statistics are selected to be of use to epidemiological researchers and public health personnel.

Instructions

- Select **ANALYSIS** from the **Programs** menu.

Information

Once **ANALYSIS** is started the screen will be divided into several distinct sections. On the top few lines of the screen is the **status information**:

```
Dataset: <None>                                Free memory: 191K
Use READ to choose a dataset
```

This status information shows the name of the current data file and the amount of free memory in the system.

The bottom line of the screen shows the function keys available from within the **ANALYSIS** module.

The rest of the screen is divided into two parts. The upper portion is labelled **Output**. This part of the screen displays the output (e.g. tables, statistics, listings) of any commands entered. The label **Screen** next to the label **Output** tells us that the output of the commands we issue will be sent to the screen. You can send output to the screen, a printer, or to a file on disk.

The lower portion of the screen is labelled **Commands**. This is where you enter commands.

Helpful Tips

The **F1** key provides help. Press **F1** at any time for help on what you are doing. The help is *context sensitive*: If you type a command and then press **F1** then **ANALYSIS** will respond with help for that particular command.

A menu of commands is available by pressing the **F2** key.

A menu of variables in the current data file is available by pressing the **F3** key.

ENTERING COMMANDS

Information

There are two methods of issuing commands to the **ANALYSIS** program. You may either type the command directly at the keyboard or choose commands and variables from menus.

Pressing the **F2** key brings up a menu of available commands. To select a command from the menu you must move the highlighted bar using the arrow keys. When the highlighted bar is over the command you wish to use press **ENTER** to select it. To execute the command press **ENTER** again. With most commands you will need to specify at least one variable.

Pressing the **F3** key brings up a menu of variables in the current dataset. Variables can be selected from the menu in the usual way. To select just one variable point to it using the arrow keys and press **ENTER**. To select a group of variables point to each one in turn and press the **+** key. When you have selected all the variables you need press **ENTER**. If you select a variable by mistake you can deselect it using the **-** key.

Getting Help on Commands

There are two ways of getting help on commands in **ANALYSIS**. If you press **F1** before you have typed a command then **ANALYSIS** will present you with a menu with options for help on general topics and specific commands. Options are selected from the menu in the usual way.

If you want help on a specific command then type the command (e.g. **freq**) and then press **F1** key **ANALYSIS** will respond with help for the command you typed.

If there is more than one screenful of help on a particular topic or command then **PgDn** will be displayed in the bottom right hand corner of the help window. Pressing **PG DN** will display the next screenful of information.

The help system has some 'hypertext' features that allow you to view help information on related topics. Related topics are shown as yellow CAPITAL letters (black on grey when selected). The first related topic is selected automatically and is shown in black on grey CAPITAL letters. You can select another topic by using the arrow keys. To view help on a selected topic press the **ENTER** key.


Press **ESC** to return to **ANALYSIS**.

CHOOSING A DATA FILE

Information

Before performing any analysis we must first tell **ANALYSIS** which data file we wish to work with (in this case **OSWEGO.REC**). The **ANALYSIS** command to retrieve a data file is **read** followed by the filename. Once you have retrieved a dataset the status lines on the screen will change to show the name of the dataset and the number of records in the dataset.

Instructions

- Enter the command **read oswego.rec** and press .

Information

The file OSWEGO.REC contains data collected from a food poisoning outbreak that followed a church supper in a small American village. This outbreak, together with details of a follow-up investigation, is described in the addendum.

Helpful Tips

For datafiles created using the **ENTER** module of EpiInfo you do not need to specify the .REC extension to the filename.

ANALYSIS can also read files created using dBase. To retrieve a dBase file you need to specify the .DBF extension to the filename.

If you can't remember the name of the file you want to retrieve then issue the **READ** command without a filename. **ANALYSIS** will respond with a menu listing all .REC files in the current directory.

EXAMINING DATA

Instructions

- Enter the command **variables**
- Enter the command **list age ill sex**
- Enter the command **freq ill**
- Enter the command **freq onsetdate**
- Enter the command **freq age**

Information

The **variables** command displays a list of variables in the current data file together with information about their type and length.

The **list** command displays the contents of the specified variables on a case by case basis.

The **freq** command displays a frequency table of the specified variable. The table lists frequencies, percentages, and cumulative percentages. If the specified variable is NUMERIC then summary statistics will also be displayed.

Helpful Tips

If the command you issue produces output that will not fit onto a single screen then **ANALYSIS** will pause at the end of each screenful of output and display the message **<more>** at the bottom of the output screen. Press any key (apart from **ESC**) which terminates the output) to see the next screen of the output.

You can scroll through the output generated by previous commands using the **PG UP** and **PG DN** keys. **PG UP** and **PG DN** move through the output one page at a time. You may also use **CTRL** + **PG UP** and **CTRL** + **PG DN** to move through the output one line at a time.

You can scroll through previously entered commands using the **↑** and **↓** keys.

The **list** command produces a simple list. **ANALYSIS** provides the **browse** **F4** command to list and browse through records in spreadsheet format. The command keys available in **browse** are shown at the bottom of the screen once you have entered the command **browse** or pressed the **F4** **browse** key. The **browse** command does not allow you to edit or add data. The **update** command allows you to edit or add data in the same spreadsheet format. A quick way to start **update** is to press the **F6** key.

It is often useful that a record of the output is kept for future reference. You can use the **route** command to send a copy of the output to a file on disk (**route filename**) or to the printer (**route printer**). Make sure that the printer is connected and 'on-line' before issuing the **route printer** command. To stop sending output to the printer or to a file issue the **route screen** command.

The **route** command will only redirect the output of subsequent commands. If you want a record of an **ANALYSIS** session then the **route** command should be the first command you issue.

TABLES AND STATISTICS

Information

ANALYSIS provides the **tables** command to produce crosstabulations. The **tables** command also produces a complete set of statistics for both simple crosstabulations and stratified analysis. We can use the **tables** command to discover which foodstuffs were responsible for the food poisoning outbreak.

Instructions

- Enter the command **tables bakedham ill**
- Enter the command **tables spinach ill**
- Enter the command **tables mashedpota ill**
- Use the **tables** command to investigate the associations between the other exposure variables and reported illness.

Information

From our output we can see that VANILLA is implicated (remember that we are looking for tables that produce a relative risk (risk ratio) greater than one together with a confidence interval that does not include one or and a large chi-square statistic together with a small p-value).

The interpretation of the Relative Risk depends on the orientation of the table. The correct format for the command is:

```
tables <exposure> <outcome>
```

ANALYSIS is helpful in that it will point out statistical measures that it considers to be significant. The program's definition of significance is somewhat rigid (i.e. $p \leq 0.05$).

Helpful Tips

An asterisk (*) can be used in the **tables** command to indicate all variables. The command **tables * ill** would produce similar output to the string of commands you entered.

Sometimes when you use the asterisk (*) **ANALYSIS** will produce tables that are both useless and very big. This is because it is treating continuous or identifying data as categorical data. You can instruct **ANALYSIS** to exclude certain variables by using the **not** keyword. The command **freq * not age** produces a tidier output than the command **freq *** on its own.

If you want to display column and row percentages in your tables then issue the command **set percents = on** before issuing **tables** commands. The command **set percents = off** will turn them off again. The values in each cell are presented as count, row percentage, column percentage. Displaying percentages can sometimes make tables difficult to read. Use the command **set lines = on** to print separating lines between the cells of tables. This will make them easier to read.

STRATIFIED ANALYSIS

Information

The `tables` command can also be used to perform stratified analysis (useful in identifying confounding factors and establishing independent associations). This technique is not needed for the analysis of the OSWEGO dataset.

Instructions

- Enter the command `tables vanilla ill sex`

Information

This command produces two tables (for male and female cases). A *Mantel-Haenszel* stratified analysis follows the tables.

ANALYSIS commands work with all records in the current dataset. You can instruct **ANALYSIS** to work with a *subset* of your data using the **select** command. The **select** command allows you to specify an expression that must be true for a record to be processed. The command `select sex = "M"` instruct **ANALYSIS** to work only with those records where the variable SEX contains the value **M**. The **select** command is a *cumulative*. Any **select** command works on the subset of data that has been previously selected. The two commands:

```
select ill = "Y"  
select age < 35
```

will select all records where the variable ILL contains the value **Y** and the AGE variable contains a value less than 35. The two commands:

```
select ill = "Y"  
select ill = "N"
```

will select NO records because the variable ILL cannot simultaneously contain the value **Y** and the value **N**.

The current selection criteria is displayed with the status information at the top of the screen. The command `set criteria = on` instructs **ANALYSIS** to print the current selection criteria before each list, frequency, or table.

You can cancel the current selection criteria and return to working with all cases in the dataset by issuing the **select** command without a condition.

The following commands produce a similar output (without the Mantel- Haenszel stratified analysis) to the `tables` command shown above:

```
select sex = "M"  
tables vanilla ill  
select  
select sex = "F"  
tables vanilla ill  
select
```

CONTINUOUS VARIABLES

Information

The **tables** command is used for data items that are arranged in categories. If you have continuous data such as unmodified measures of height, weight or age then you will need to use the **means** command.

The **means** command takes two variables - the variable containing data to be analyzed and the variable indicating how groups will be distinguished.

Instructions

- Enter the command **means age ill**

Information

The output of this command produces a full set of descriptive statistics for the two groups (ILL = "Y" and ILL = "N") as well as ANOVA and equivalent non-parametric tests.

Helpful Tips

The **means** command produces a table of values (similar to a frequency table) before displaying statistics. Often this table will not be of use. You can instruct **ANALYSIS** not to produce tables by adding **/n** to the end of the command to indicate 'no tables' (e.g. **means age ill /n**).

ANALYSIS also has a command that performs multiple linear regression. The format for the command is:

```
regress <dependent> = <independent1> ... <independent15>
```

You can use up to 15 independent variables with the **regress** command.

You may also use the **means** command to produce full descriptive statistics for a single (*ungrouped*) variable:

```
means age /n
```

WORKING WITH VARIABLES

Information

In the OSWEGO dataset there is no variable telling us the incubation period for the outbreak under investigation. We need to create a new variable that will contain the incubation period for each case. The **define** command is used to create a new variable.

The **define** command when used on its own creates a new variable of the NUMERIC type. To create other variable types use the same special characters that define variable type and length in the questionnaire. The command **define temp ###.##** creates a variable of the fixed decimal type with the name TEMP.

Once we have defined a new variable INCUB we need to instruct **ANALYSIS** to calculate the incubation period from the variables ONSETTIME (time of onset) and TIMESUPPER (time of supper). Because the outbreak was spread over two days we need to correct the value of the INCUB variable to take account of this. To do this we must add 2400 (for 24 hours) to the INCUB variable for each case where ONSETDATE is 04/19 (i.e. the second day of the outbreak).

Instructions

- Enter the following commands:

```
define incub
incub = onsettime - timesupper
if onsetdate = "04/19" then incub = incub + 2400
freq incub
```

Information

The **freq** command displays a frequency table of the INCUB variable. Examination of the frequency table and the summary statistics below the table show us that the median incubation period was four hours and that the most common incubation period (mode) was three hours. The difference in the number of valid INCUB values and number of cases where ILL = "Y" is due to missing values. The odd values (e.g. 570) are due to the fact that we are performing arithmetic in base 10 instead of a combination of base 24 and base 60. One solution to this problem is to strip away the 'minutes' part (i.e. the final two digits) of the INCUB variable using the **div** operator:

```
incub = incub div 100
freq incub
```

For most purposes this approximation is sufficient. It is possible to convert times to 'decimal' times using the following *general* form of commands:

```
define dectime ##.#
dectime = time div 100 + ((time / 100 - time div 100) / 0.6)
```

You might like to repeat the calculations for incubation period using decimal times (you will need to create variables to hold the decimal equivalents of both the TIMESUPPER and ONSETTIME variables).

PRODUCING GRAPHS

Instructions

- Enter the command **histogram incub**

Information

The **histogram** command provides a graphical representation of the **freq** command. This is not a traditional *epidemic curve*. We are plotting the incubation period instead of time of onset. A traditional epidemic curve (plotting time of onset) would be produced by the following set of commands:

```
define newtime
newtime = onsettime
if onsetdate = "04/19" then newtime = newtime + 2400
histogram newtime
```

Helpful Tips

To print the graph enter the command **route printer** or press **F5 Printer On** before the **histogram** (or **pie**, **bar**, **line**, and **scatter**) command. When printing finishes enter the command **route screen** or press **F5 Printer Off**. Make sure the printer is connected and 'on-line' before attempting to print a graph. If you want to send the output of a graphics command to a file you should specify a different file for each graph. You must also use **set graph** to specify the way the graph will be saved:

```
set graph = $pcx
route incub.pcx
histogram incub
```

You can export graphs in several different formats. The most useful are: **\$pcx** = PC PaintBrush (640*480), **\$bmp** = BMP (Windows 3.x) format, **\$cgm** = Computer Graphics Metafile, and **\$wpg** = WordPerfect Graphic.

ANALYSIS also has a command that performs multiple linear regression. You can use this command to investigate the relationship between incubation period and age:

```
regress incub = age
```

Use the **scatter** command to see the regression line:

```
scatter age incub /r
```

The **/r** instructs **ANALYSIS** to calculate and plot a regression line. Note that the order of the variables is different for the **regress** and **scatter** commands.

The variables you created (INCUB and NEWTIME) will not be saved when you leave **ANALYSIS**. If you want to save variables defined during an **ANALYSIS** session then use the **route** and **write recfile** commands to create a new data file:

```
route oswego2.rec
write recfile
```


PROGRAMMING ANALYSIS

Information

So far we have used the **ANALYSIS** program in interactive mode. We have been entering commands one at a time and looking at the results from each command as it is carried out. **ANALYSIS** allows us to place the commands we wish to use in a *program file* so that they can be executed by issuing one command. A list of instructions held in a file is called a program. Programs are useful for routine analysis (e.g. for a surveillance system) or complex coding operations. The sample program repeats the analysis of the OSWEGO dataset. Output is sent to a file called **STARTING.OUT**.

Instructions

- Quit **ANALYSIS** and select **EPED** from the **Programs** menu. Set up **EPED** in TXT mode. Type the sample program shown on the next page.
- Check your work carefully and save the program as **STARTING.PGM** and return to the EpiInfo top-level menu.
- Select **ANALYSIS** from the **Programs** menu.
- Enter the command **run starting.pgm**. When the program finishes you will be returned to the EpiInfo top-level menu.
- Select **EPED** from the **Programs** menu.
- Select the **File** menu by pressing **[F2]** and select the **Open file this window** option. Type the filename **STARTING.OUT** and press **[ENTER]**. The output from the program appears on the screen.
- Check the output from the program and return to the EpiInfo main menu by pressing **[F10]**.
- Return to DOS by pressing **[F10]**.

Helpful Tips

ANALYSIS provides an easy way of developing programs. First enter the commands interactively to check if they actually produce the results you want. After you have entered all the commands use the **save** command to save them in a text file. You can then use **EPED** to edit the file. Remember that **ANALYSIS** program files should have the extension .PGM. The **save** command usually saves the last 20 commands entered. If you want to save more commands you should enter the command **set cmdstack = nn** where **nn** is the number of commands you wish to save. You should use **set cmdstack** before entering any of the commands you wish to save. The following command will save previously entered commands to a file called **NEWPROG.PGM**:

```
save newprog.pgm
```

SAMPLE PROGRAM

```
route starting.out
read oswego.rec
freq ill
define incub
incub = onsettime - timesupper
if onsetdate = "04/19" then incub = incub + 2400
freq incub
tables * ill
route screen
quit
```

GOOD LUCK

You have used some of the features of the EpiInfo package. There are some more advanced features which are discussed in the manual. The way to learn to use a statistical package is simply to practice and experiment. You might like to run through this tutorial again to boost your confidence. The EpiInfo package comes with several example datasets and worked examples can be found in the manual.

OSWEGO COUNTY REVISITED

The following document is taken from Public Health Reports, Vol 91, No. 2, March - April 1976. The author, Dr Michael Gross is medical director, Valley Community Health Centre, Picture Rocks, and assistant director, Family Practice Residency, Williamsport Hospital, both in Lycoming County, Pa. At the time of writing he was Public Health Service Epidemic Intelligence Service Officer in the New York State Department of Health, Albany.

Oswego County Revisited

A review of documents and interviews with surviving principals of a famous 1940 outbreak of staphylococcal gastroenteritis.

After a church supper in the small village of Lycoming, Oswego County, New York, on April 18th, 1940, a majority of participants became ill with gastroenteritis. Epidemic investigation by the New York State department of Health suggested that the agent was *Staphylococcus Aureus*, with home-made vanilla ice cream implicated epidemiologically by food histories and culture as the vehicle of infection.

Although the source of contamination of the ice cream was not established at the time, the appealing story and elegant epidemiology make the data ideal for a teaching exercise and this is why the outbreak has become famous.

The outbreak was initially used as an epidemic example in 1940 by Wendell Ames, MD, then New York State Epidemiologist, at the Albany Medical College. In 1942 it was taken from New York State to the US Navy Epidemiological Training Program in Bethesda, Md., by Stafford Wheeler, MD, and used for training by him and Alexander Langmuir, MD, during World War II. In 1946 Langmuir brought it to the John Hopkins School of Hygiene and Public Health, where it was used as a laboratory exercise in the basic course in general epidemiology. In 1949 Langmuir joined the newly established Epidemiology Program of the Communicable Disease Centre (now the Center for Disease Control), and what has become known as the 'Oswego Problem' has been used in the Epidemic Intelligence Service training course since its first class in 1951. Use of this training exercise has been spread throughout the United States and to several other countries by the generation of persons exposed to it in their training, and it is now widely used in epidemiology training courses conducted by various health departments. Figures from this exercise are also presented in two manuals devoted to the investigation of foodborne disease outbreaks; these manuals have become important teaching resources (1, 2).

The Investigation, 1940

Records of the original outbreak investigation are preserved in memorandums and publications of the New York State Department of Health (3, 4, 5). Most of the cases occurred during the night of the supper, April 18th, and the village health officer reported the epidemic to the State Health Department the next day. The district health officer went to the scene immediately to verify the existence of the outbreak, and a team, including two epidemiological trainees, was assigned to make a full investigation, including food and illness histories, on April 20th. Their report states:

“All handlers of the ice cream were examined. No external lesions or upper respiratory tract infections were noted” (5)

Nose and throat cultures of the two who made the ice cream mix, however, did grow *S. Aureus*.

The ingredients used to prepare the ice cream were fresh, unpasteurized, milk, condensed milk, sugar, eggs, flour, chocolate, and vanilla extract. The chocolate and vanilla mixes were prepared in separate containers on April 17th, the day before the supper, and left covered overnight before being frozen. The raw milk was an unlikely source of infection, because the chocolate ice cream was not contaminated. No plausible explanation for contamination of the vanilla ice cream was obtained, despite intensive efforts during the several weeks following the outbreak. Lycoming is a very small rural village located two miles from the south east shore of Lake Ontario. At the time of this outbreak the area was best noted for its production of strawberries; the economy is now based on lettuce farming with substantial contributions from a nuclear power station on the lake shore. On a map (6) published in 1954, only 42 dwellings could be counted in the immediate area.

Follow up Investigation 1972

I visited Lycoming in March 1972 and interviewed the surviving principals of the church supper. Additional evidence obtained at the time suggested that one of the food handlers may, in fact, have had an external lesion on her hand when she made the ice cream. The part-time village health officer who initially reported the outbreak was still in practice; he recalled the outbreak vividly as a major event in his career. The morning after the supper he made a routine home visit to an invalid patient. At the time one of the makers of the ice cream, employed part-time in the same household, showed him a small draining infection on one of her fingers. She was not ill, no culture was taken, and no written record was made. Obviously if such a lesion had discharged while she was stirring the mix or wiping the spoon, that batch of ice cream could easily have been contaminated, and overnight storage at room temperature would have allowed multiplication of the organism and production of toxin.

Attempts were made to assess the plausibility of this revelation. There is no detailed description in the records of the New York State Health department of the examination of food handlers, but only the general statement of negative results cited previously. The two primary makers of the ice cream were still alive, but neither recalled such a lesion. Their examination by the epidemiological investigators presumably took place late on April 20th - 2 days after the supper and 3 days after the ice cream was made. Three days allow ample time for the resolution of uncomplicated staphylococcal skin lesions. The village health officer examined the participants on the morning of April 19th.

From reports of the personality of the woman who had the lesion, she would not have volunteered this significant historical information to an outsider. She was hospitalised for an unrelated condition from April 21st to May 5th, 1940. Records of that hospitalisation also fail to confirm traces of the lesion. Thus, she would not have been available for the additional questioning by the investigators from the State Health Department which several villagers remember from this period.

Perhaps most remarkable is why the evidence presented here remained in Oswego County for almost 33 years. 'Village Health Officer' in New York State was, and is, a part-time position with nominal reimbursement. These officials are required to report the existence of outbreaks, but they are not responsible for detailed investigation. The village health officer apparently was not interviewed. Was it because of the excitement or because each investigator thought the other had interviewed him? Neither of the two former epidemiological trainees remembers meeting the village

health officer during the investigation. They also cannot specifically remember examining the food handlers or any other details which might contradict the hypothesis presented here. It is symbolic of the loneliness of the field investigator that both now recall doing the entire job alone.

An annual church supper is still held in Lycoming, and home-made ice cream is still the chief attraction. The two surviving food handlers of 1940, both vigorous in their 70's still help out.



Although the installation of refrigeration in the early 1950s put an end to overnight incubation of the ice cream mix, they readily admit that their lapse in proper food handling continued until then. But a second epidemic did not occur.

References

1. Bryan, FL : *Guide to investigating foodborne disease outbreaks and analyzing surveillance data*. CDC, Atlanta 1973.
2. Committee on Communicable Disease Affecting Man, International Association of Milk, Food, and Environmental Sanitarians: *Procedure for investigating foodborne disease outbreaks*. Shelbyville 1966.
3. State of New York: *61st annual report of the Department of Health for year ending December 31, 1940*. Albany 1940.
4. Gastroenteritis in Oswego County traced to home made ice cream. *Health News* 17:104 (NYSDH) 1940
5. Memo. Gastroenteritis outbreak, Lycoming. *Final Report Bureau of Epidemiology, Albany*. May 24, 1940.
6. US Geological Survey: Oswego East quadrangle. Washington DC, 1954.

INDEX

*.CHK files	6, 29	Cancel command, in EPED	4
*.PGM files	6	Case, definition of	15
*.QES files	6	Centre text, in EPED	4
*.REC files	15	CMDSTACK	37
*.DBF files	30	Coding data	22
ANALYSIS program	28	Commands, in ANALYSIS	29
ANALYSIS	32	Control characters in EPED	13
BAR command	36	Copying text, in EPED	11, 12
Continuous variables	34	Correcting text, in EPED	4, 7
dBase III+ files	30	Cross-case data checks	26
Defining variables	35	Cursor movement, in EPED	7
Entering commands	29	Data, editing	24
FREQ command	31	Data, checking (batch)	16, 25, 26
Function of	16	Data, checking (interactive)	16, 25, 26, 27
Graphing commands	36	Data, checking	25, 26, 27
HISTOGRAM command	36	Data, coding	22
Help	29, 30	Data, double entry	16, 26
IF command	35	Data, entry	23
LINE command	36	Data, errors in	25
LIST command	31	Data file, definition of	15
MEANS command	34	dBASE III+ files	30
Menus	29, 30	Deleting text, in EPED	11, 12
NOT keyword	32	Designing questionnaires	18, 19, 22
Output	32	Double entry	26
PIE command	36	ENTER program	23, 24
REGRESS command	34, 36	ENTER	
Retrieving files	30	Changing file format	19
ROUTE command	31, 36	Function of	16
SAVE command	37	Editing features	24
Saving files	36	EPED	
SCATTER command	36	Block operations	11, 12, 13
Screen described	28	Control characters	13
SET CMDSTACK	37	Cancel command	4
TABLES command	32, 33	Centre text	4
VARIABLES command	33	Copying text	11, 12
BAR , command in ANALYSIS	36	Correcting text	4, 7
Batch data checking	16, 25, 26	Cursor movement	7
Block operations in EPED	11, 12, 13	Deleting text	11, 12
CHECK program	26	Document format	4
CHECK		Finding text	9
Changing variable names	27	Formating block	12
Setting up checks	27	Format paragraph	4
Function of	16	Function and uses	1, 2, 3
Types of data checking	26	Help	3
.CHK files	29	Indent	3

Insert mode	2, 3	NOT , keyword in ANALYSIS	32
Justification	3	OSWEGO.REC data file	1
EPED		Output, in ANALYSIS	32
Menus	4	Overtime mode, in EPED	3
Moving text	11, 12	PIE , command in ANALYSIS	36
Overtime mode	3	Power cuts	6
Printing text	6, 13	Printing text, in EPED	6, 13
Repeat find/replace	9, 10	Questionnaire, example	20
Replacing text	10	Questionnaire, design	18, 19, 22
Retrieving files	7	REGRESS , command in ANALYSIS	34, 36
SET menu	3	Repeat find/replace, in EPED	10
Saving text	6	Replacing text, in EPED	9, 10
Screen described	2	Retrieving files, in ANALYSIS	30
Search direction	9, 10	Retrieving files, in EPED	7
Search options	9, 10	ROUTE , command in ANALYSIS	31
Setting up	3	SAVE , command in ANALYSIS	37
Starting	2	Saving text, in EPED	6
Typstyles	13	Search direction, in EPED	9, 10
Variable definition	19	Search options, in EPED	9, 10
WordStar commands	3	Setting up EPED	3
WordWrap	2, 3	Starting EPED	2
EpiInfo		Starting EPIINFO	1
Description	1	Stratified analysis	33
Filenames	6	Survey processing	16
Starting	1	TABLES , command in ANALYSIS	32, 33
Survey processing	16	Typstyles, in ANALYSIS	32, 33
Epidemic curve	36	VALIDATE , function of	16
Errors in data	25	VARIABLES , command in ANALYSIS	33
FREQ , command in ANALYSIS	31	Variables	
Filenames, EpiInfo	6	AUTO-INCREMENT	18
Filenames, MSDOS	6	Changing names, in CHECK	27
Finding text, in EPED	9	Continuous	34
Format paragraph, in EPED	4	DATE	18, 22
Graphics	36	Defining in ANALYSIS	35
Help  , in ANALYSIS	29, 30	LAST-UPDATE	18
Help  , in EPED	4	LOGICAL	18, 22
HISTOGRAM , command in ANALYSIS	35	Length	17, 22
IF , command in ANALYSIS	35	NUMERIC	18, 22
Indent, in EPED	3	Name	17, 21
Insert mode, in EPED	2, 3	PHONENUM	18
Interactive data checking	16, 25-27	TEXT	18, 22
Justification, in EPED	3	Type	17, 18, 22
LINE , command in ANALYSIS	36	YES/NO	18, 22
LIST , command in ANALYSIS	31	Attributes of	17
MEANS , command in ANALYSIS	34	Defining	19, 35
Main menu, selection	2	Defining from EPED	19
Menus in ANALYSIS	28, 29	In ANALYSIS	31
Menus in EPED	4	WordStar commands, in EPED	3
Moving text, in EPED	11, 12	WordWrap	2, 3

