

I Ching Squared v2.1

by Eli Meir

Copyright © 1993 by Eli Meir

| | |
|-------------------------------------------------|----|
| Introduction..... | 1 |
| Sample Run Through..... | 2 |
| Files..... | 5 |
| Input File Formats..... | 5 |
| Output File Format..... | 5 |
| Selecting Files for Processing..... | 6 |
| Filename Templates..... | 7 |
| Selecting Files for Phases and Differences..... | 9 |
| Processing..... | 11 |
| Event Lengths..... | 11 |
| Periods..... | 11 |
| Differences..... | 12 |
| Phases..... | 12 |
| Original Phases..... | 13 |
| One-to-One Phases..... | 13 |
| Lockstep Phases..... | 14 |
| Clean Phases..... | 15 |
| Histograms..... | 17 |
| Filtering Files..... | 17 |
| Finding Statistics Without Processing..... | 17 |

| | |
|----------------------------------|----|
| Statistics..... | 18 |
| Selecting Statistics..... | 18 |
| Filename..... | 18 |
| Type..... | 19 |
| Number of Points..... | 19 |
| Mean..... | 19 |
| Standard Deviation..... | 20 |
| Standard Error..... | 20 |
| Coefficient of Variation..... | 20 |
| Median..... | 21 |
| Quartiles (Q-1/4 and Q+1/4)..... | 21 |
| Chi Square..... | 22 |
| Sum..... | 22 |
| Sum Squares..... | 22 |
| Cut and Cut Side..... | 22 |
| Options..... | 23 |
| Operational Unit..... | 23 |
| Data Type..... | 24 |
| Save Options..... | 24 |
| Phase Type..... | 25 |
| Cutting..... | 25 |
| Saving the Options..... | 25 |

| | |
|----------------------------------------------------------|--------|
| Word Processor Functions..... | 26 |
| New Statistics File..... | 26 |
| Saving Statistics Files..... | 26 |
| Opening Files..... | 26 |
| Why is the Font Courier 10 Included on the Disk?..... | 26 |
| Help..... | 27 |
| Programmers Appendix..... | 28 |
| Adding a Processing Command..... | 29 |
| Adding a New Statistic..... | 30 |
| Selected Pseudo-Code..... | 30 |
| Phase Mean | 30 |
| Phase Standard Deviation | 30 |
| Phase Median | 31 |

Introduction

I Ching Squared is a straightforward statistics program designed to analyze the event onset/offset files (event files) produced by the program Spike Studio. Currently, I Ching can determine event lengths and periods between events for individual recordings. I Ching can also measure differences (or phases) between events from two recordings. Histograms can be made from any type of data, including those mentioned above. I Ching stores the numbers from these analyses in text files, along with other user selected statistics, including means, standard deviations, and chi squares. These files can be examined using a word processing program, spreadsheet, or graphing program.

In addition, I Ching will perform the same operation on multiple files with a single command. Thus if you have many onset/offset files from one experiment and would like to analyze them all in the same way, using I Ching you can analyze all the files at once with one command.

This manual is divided into several sections. The first section will give step-by-step instructions for a sample analysis session. Next are sections dealing with each part of I Ching specifically.

This manual assumes familiarity with the Macintosh. If you've never used a Mac before, read an introduction to the Macintosh before you read this manual. In the manual, the input files are assumed to be the event files produced by the Spike Studio program. I Ching should work for files from other places too, but check the File Format section to be sure.

I Ching is constructed so that it is easy to add new processing functions or statistics to the existing ones. The author is willing to add in additional functions to the program, and if you wish to modify it yourself the source code is available. This manual does not discuss the source code for I Ching, though there is a small programmers appendix at the end of this manual. In general however, it is probably easier and faster to have the author modify the program for you.

I hope I Ching is useful for analyzing your data. If you have questions, comments, bug reports, or want help write or email:

Eli Meir
Spike Studio
149 Grandview Ct.
Ithaca, NY 14850
meir@zoology.washington.edu

Sample Run Through

The foundation of I Ching Squared is a simple word processor, so you should start by pretending that you are using a typical, fairly simple-minded Macintosh word processor, where the program every now and then decides to throw a line of text into the document.

Run the program. A copyright message will appear which you can get rid of by clicking or hitting a key. Then a new window will appear with one line of writing at the top. This window is like the document in a word-processor, and its where all the statistics will go.

Let's suppose you want to find the mean, standard deviation and median of the periods between events in one of your data files (the period is the time from one event to the next). Select 'Set Statistics' from the 'Statistics' menu. A dialog box will appear with all the statistics included in the program. Type in a 1 next to Filename, a 2 next to Type, a 3 next to Num Points, a 4 next to Mean, a 5 next to standard deviation, and a 6 next to median. Put 0's next to all the other stats. This indicates that when you process the file for periods, you would like a line of text to appear with the filename, the fact that you processed it for periods, the number of periods found, the mean, standard deviation and median, and that they should appear in that order.

Now simply go to the 'Statistics' menu and select 'Get Period Stats'. A normal macintosh file dialog box will pop up and you should select the name of the file you want. The program will process this for a few seconds, and then a new line will appear in the window, containing all of the statistics that you selected above.

Now suppose you now want to save all the periods from that file in a new file (perhaps so that you can view them in a spreadsheet). Go to the 'Options' menu and select the 'Save Options' item. A dialog box will ask which types of processing you would like saved. Check the 'Periods' box and click 'OK'. Now select 'Get Period Stats' again. A dialog box will appear, where you should select the data file again. Then another dialog box will come up, requesting the name of the file in which you want to save the periods. When you enter the name, I Ching will spend a few seconds processing, and then a new line will appear in the window with the name of the file where the periods were saved, and the statistics for those periods.

Notice that you now have two lines with period statistics for the same file, one from when you just got statistics and one from when you saved the periods (unless you were sneaky and picked different files) This is rather ugly, and you probably want to get rid of one of the lines. No problem. Remember, this is a word processor, so just like in any word processor, you can select the line you want to get rid of and 'Cut' it, or just press the delete key. Presto, the offending characters are gone. You can also change the font, the font size, move the lines around, add comments, or even edit your statistics if you don't like them (just kidding - I'm sure you're too honest for that).

Now that you're getting comfortable with the program, let's suppose you want to get periods for the whole experiment. This is as easy as getting them for just one file, but first you must make sure that I Ching is set up properly for your file naming conventions. If someone has already set up the program for your laboratory, you can probably skip ahead to the next paragraph. However if you're the first user, you will need to set the filename template that the program uses. I Ching knows that certain files come from the same experiment because you tell it what a filename should look like, and part of the filename identifies the experiment. This is a little bit complex, so rather than repeating it here look ahead to the Filename Template section of the manual (Section 3-4), set up a template for your filenames, and then come back here and continue with the sample run-through.

Now that you have the filename template set correctly, go to the 'Options' menu and select 'Experiment'. Now 'Get Period Stats'. A dialog box will prompt you for one of the files from the experiment. It doesn't matter which file you pick, as long as its from the experiment you wish to analyze. Next another dialog box will appear asking where the period files should be saved (remember that since the save periods option is still checked, a period file will be saved for every event file in the experiment). The only thing that matters in the name you choose is the part that will replace the experiment identifier in the filename (the part that is not between two *'s). This part will be used as the experiment identifier in the output filenames, with the the rest of each filename copied from the input filenames. I Ching will automatically give you a name it thinks is appropriate, in this case the experiment identifier with a "per" added after it, and if this name is acceptable then you can simply click OK (Section 3-3).

Now just sit back and wait. As each file is processed a period file will be put on the disk and the period statistics will be displayed on the screen. If you have selected a long experiment, you can go get a cup of coffee. When the program thinks it has reached the end of the data files for this experiment, a dialog box will appear asking if the program should go on. If you have skipped a condition in you experiment (maybe you have conditions 1, 2, 3, and 5), the program will stop after 3, because it doesn't find files in condition 4. In this case, you would answer yes, and the program would continue searching for files in the next condition. If all the data for that experiment has been analyzed, select no.

If you want to save the file with all the statistics (that you are now viewing), go to the 'File' menu and select the 'Save' item, the same way you would if using a word-processor. Similarly, if you want to print the statistics, select 'Print'. If you want to look at one of the period files you just created, select 'Open' and open the period file. You should consider all these windows to be just like document windows in a word-processor.

Finding event lengths (the time between the onset and offset of an event) works the same way as finding periods. Phases and differences work a little differently. Suppose you want to get the phases between roots 1 and 2 in all the conditions of your experiment. 'Experiment' in the 'Options' menu should still be checked from before, so just select 'Get Phase Stats' from the 'Statistics' menu. A dialog box will pop up with a large blank area and a few buttons on the bottom. Click on the 'Add Phase' button. Another dialog box will come up with two or four boxes. These boxes allow you to type in the label and root number of the lead and following data files which you want to get phases from. Right now, instead of typing in the labels and/or the roots, click on the 'Pick Files' button. A standard file requester will appear, and you should pick a file which you would like to be the lead file in a phase calculation. Then another standard file requester will appear, and you should pick a file which you would like to be a following root in the phase calculation.

Now the label and root number fields in the Add Phase dialog box should be filled in appropriately, so you can click 'OK'. A new entry will appear in the original dialog box, which shows the identifiers for the files you just picked. The set of identifiers before the dash indicates the lead files, and the identifiers after the dash are used to find the following files. For every condition in the experiment, I Ching will look for a filename with the label and root of the lead file and another file with the label and root of the following file, and will find the phases between these.

Now a dialog box will appear, where you should pick any file from the experiment. This file is just going to be used to get the experiment identifier. Then go get another cup of coffee while the program churns away and spits statistics onto the screen.

As a general rule: whenever you are getting some sort of statistic for a single file, you should put that specific file into the first file requester that comes up. When you are getting statistics for a condition, you can pick any file from that condition. When you are getting statistics for the whole experiment, you can pick any file from the experiment.

If you are saving the processed information as well as getting statistics on it, then a second dialog box will always appear where you must enter the filename where you want the information to be saved. If you are getting statistics for just a single file, then the name that you enter is the name where the info will be saved. If you are getting statistics for a condition or a whole experiment then the only part of the output filename that you type in will be the experiment identifier. The rest of the output filenames will be copied from the input filenames.

Files

I Ching works exclusively with text files. Thus all data which you would like I Ching to make calculations on must be saved as text. It can, however, be saved in a variety of formats (see below). When I Ching outputs data, it uses a single format and precision for almost all types of data. These input and output formats are explained in Sections 3-1 and 3-2 below.

The names of the files must follow a certain format if you want to process more than one file at a time. Sections 3-3, 3-4, and 3-5 describe how to select files to be processed, and discuss how you can set the filename formats.

Input File Formats

I Ching normally operates on onset/offset files, which contain the onsets and offsets of consecutive events (bursts or spikes, for instance). I Ching does not require a specific format for these files. What it does need is some non-numeric character or characters separating each onset from its corresponding offset, and there must be an offset for each onset (an even number of numbers). For most of the processing functions, either the onset, offset, or midpoint of each event will be used. The Data Type function is used to decide between these (Section 6-2). I Ching can also operate on files which are not onset/offset files. In this case, you should indicate that each point is to be used, or that the file contains processed data (Section 6-2).

Output File Format

The processed output files created by I Ching all use the same format. They contain a single column of numbers, with the first number being the value calculated from the first data point(s), the next number being the value calculated from the next data point(s), etc.. Each number is saved as text, with 3 digits of precision after the decimal point. The numbers are right justified. The only exception to this is when time stamps are being saved, in which case there are two columns of numbers, with the first being the time stamps and the second being the data values (Section 6-3).

Selecting Files for Processing

The selection of files for processing is done in a standard way throughout the program. When you request a type of processing, a dialog box will appear requesting the input file. Then if you asked that the processing be saved, a save file requester will prompt you for the name of the output file. The exception to this are phase and difference processing, which are handled slightly differently from all the other processing and are explained below (Section 3-5).

If you are processing one file at a time, you will select each file that will be used. If you are processing by condition or experiment, then the filenames which you gave in the requester boxes will be used as templates. Filename templates, which are explained below, are what allow I Ching to process multiple files with one command. I Ching will look for and process all the files in the current folder which are from the same experiment as the file you selected. Specifically, I Ching will find all the files in the current folder with the same experiment name (and condition if you're processing by condition) as the file you selected, process all of those files, and save them with the experiment name of the output file you typed in, but with the label, root, and condition of the input file. The current folder is the one which contains the input file you selected; all the output files will be saved in the folder which you selected in the save file dialog box.

Filename Templates

The most powerful feature of I Ching is its ability to process multiple data files with one command. I Ching uses the filenames to figure out which files the user has requested for processing. In order for I Ching to use the filenames, they must all follow a certain format, and this section of the manual outlines the possible formats, how they are used by I Ching, and how to change the file format.

I Ching can process files from either a condition of an experiment or a whole experiment. The condition and the experiment must be specified in the filename. In addition, a character and a number, called the label and the root, can be specified in the filename. When I Ching is looking for all the files in a condition, it will search for all files with the experiment name and condition number that have one of the possible labels and roots. You can design a format to put the experiment name, condition number, label, and root in almost any order you want, and give whatever meaning you want to each of them.

The condition is a number which should start as 1 for the first set of files from the experiment, and then increase by one for each successive set. If there is no natural way to break up the files from your experiments into chunks which you would want to process together, then you may decide to have just one condition per experiment. Note, however, you must always have a condition number in the filename.

The experiment name is called the body of the filename. Every file must have a body, and this should somehow specify the experiment from which the files came. The date of the experiment makes a logical and easy to use body.

The label can be any one alphanumeric character placed somewhere in the filename. Possible uses are to differentiate between recordings from different positions on the preparation (e.g., L for left, R for right), or different types of recordings. You specify which characters can be used as a label, and when searching for all files in a condition, I Ching will look in turn for files containing each of these labels. It is not necessary to have a label in the filename template.

The root can be any number within a range of numbers which you specify. It might be used in conjunction with the label, in order to identify different positions on the preparation. You must specify the range of numbers which the root is allowed to take; when searching for all files in a condition, I Ching will look in turn for files containing each of the possible root numbers. It is not necessary to have a root number in the filename templates.

You may also have other characters in the filenames, but they must always be present and they must always appear in the same part of the filename.

You may combine all of these in almost any way you want to come up with a file format. Once you have decided on the format, you must tell it to the program. To do this, use the 'Change Template' function in the 'File' menu. When you select this item, a dialog box will appear in which you can enter the template for your file format. At the top of the dialog box is a text field where the current template is shown. The parts of the template are given in the text field as follows:

| | |
|--------|-------------------------------------------------------------|
| *lab* | The label part of the filename |
| *root* | The root part of the filename |
| *body* | The part of the filename which identifies the experiment |
| *cond* | The condition number of the filename |

You may change the positions of these by changing the format in the dialog box. A few examples are given below.

Below the template are a few parameters for some of the fields. There are two boxes where you should type in the minimum and maximum possible root numbers. All root numbers from minimum to maximum (inclusive) will be valid root numbers. There is also a box where you can type in all the characters which can be valid labels. Type the labels one after the other, with no spaces inbetween. Uppercase and lowercase letters are the same. There is a maximum of 70 labels allowed. (In some versions of the program there may be a 'Special Body' check box. This box will be present if you name your files in a way that cannot be used as a normal template by the program. In this case, it is possible for special code to be added to I Ching to deal with your specific template, and this special code will be executed if the 'Special Body' box is checked.)

There are some restrictions on the arrangement of different items in the template. *There must always be a body and condition number*, but the label and root are optional. Since both the root and the condition are numbers, there must be something separating them which is not a number (common separators are periods, slashes, backslashes, etc.). If the label might begin with a number and it comes after the root or condition, then there must be a separator between the root or condition and the label. There must also be a character which denotes the end of the body, unless the body is the last thing in the filename. The character which denotes the end of the body cannot appear in the body itself.

Here are some examples of different filenames, and what their templates would look like.

R17.19/04/91.1 *lab**root*.*body*.*cond*
DP(5).tuesday 19-08.7 DP(*root*).*body*.*cond*
5HT(8)-10min/T76/4
body/*label**root*/*cond*
space_shuttle-2 *body*-*cond*

Selecting Files for Phases and Differences

If you are processing phases or differences one file at a time, then two file requesters will appear instead of the usual one. The first file requester is where you select the file you would like to be the lead file, and the second file requester is used to select the following file.

If you are processing for phases or differences for a condition or experiment, there is an extra step in the procedure. You must specify which pairs of files in each condition you would like processed for phases or conditions. The program includes a set of dialog boxes which allow you to specify up to 18 pairs of files, or to process every pair of files in each condition. You can enter the pairs using the 'Set Roots for Phases' item in the 'Options' menu. Whenever you select 'Get Phases' or 'Get Differences' for a condition or experiment, you will also have an option to change the pairs of files. In either case, a dialog box will appear with a large area in the middle for the pairs of files, and an 'Add Pair', 'Delete Pair', and 'Change Pair' buttons, as well as an 'All Pairs' check box.

The pairs will be shown in the middle of the dialog box. A label and/or root number will be shown to identify each file of each pair, separated by a dash. The identifiers on the left of the dash show what will be the lead file, and the identifiers on the right of the dash show what will be the following file. For each condition which is being processed, I Ching will look for files with the same identifiers as the identifiers in each pair, and will calculate phases or differences for these pairs of files.

To add pairs of identifiers, use the 'Add Pair' button. When you click on this button, a second dialog box will come up, with two or four boxes in it where the labels and roots can be entered. The top boxes are for the lead file, and the bottom boxes are for the following file. If you do not have a label or do not have a root number in your filenames, then there will only be one box on top and one on the bottom, otherwise there will be two and two. You can type the labels and root numbers directly into these boxes, or you can pick actual files and let the program extract the label and root numbers. To pick files, click on the 'Pick Files' button. Two standard macintosh dialog boxes will appear; in the first one you should select a file which has the identifiers of the lead root, and in the second one you should select a file which has the identifiers of the following root. The appropriate labels and roots will then be placed in the boxes of the 'Add Pair' dialog, and you can click OK to enter this pair.

To delete a pair that you no longer want to use, click on the 'Delete Pair' button, and then click on the pair that you want to delete. The pair will be deleted, and all the pairs after it will be shifted up one box.

If you want to change a pair, click on the 'Change Pair' button, and then click on the pair that you want to change. The second dialog box will now appear, and the label and root boxes will contain the labels and roots from the pair you wanted to change. Make the appropriate changes and click OK.

There may be times when you want to get the phases or differences between every possible combination of the files in a condition. In this case, instead of entering each pair one at a time, you can simply click on the 'All Pairs' check box. When 'All Pairs' is checked and you process a condition or experiment for phases or differences, I Ching will find all the files for each condition and process every possible pair of files in the way specified. The lead files will always be the files with the lower root number or the label which comes earlier in the string of labels which you typed in for the filename template.

Once you have picked the pairs you would like to have processed, the procedure is the same as previously described for all other types of processing.

Processing

In I Ching, processing refers to reading an input file, usually an onset/offset file, and calculating a new set of data from the input data. Possible types of calculations currently include finding event lengths, periods, phases, differences, and histograms. There are various options in the type of data to use and how to do the calculations, and these options are discussed separately in the options section (Section 6). Statistical analyses can then be performed on this newly calculated data set. Statistics are discussed separately in the statistics section (Section 5). This section discusses all the types of processing that are currently included in the program.

Many of the following processing functions discuss comparisons between data values. A data value can be either the onset, offset, or midpoint of a event as decided by the user (see Section 6-2).

Event Lengths

Event lengths are the times between the onset and the offset of each event. They are found by subtracting the onset time from the offset time for every event.

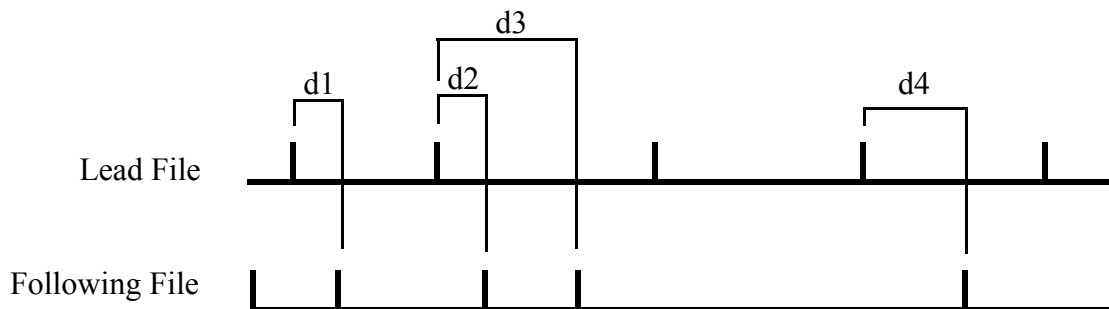
Periods

Periods are the time between each data value and the next data value. The first period is found by subtracting the first data value from the second, the second period is the third data value minus the second, and so on.

Differences

The difference is the time between a data value in one file and the next data value in a second file. Differences are related to the phase between the data points in the files. They give an indication of how delayed the data points in the second file are relative to the first. If the differences between the two files are relatively constant, this may indicate some coordination between the systems which generated the two data sets.

Differences are found using a lead and a following data file. The first and second data values in the lead file are found (call them L1 and L2). The data values in the following file are then scanned until one appears which is between the first and second values in the lead file (call this F). The difference equals $(F - L1)$. If the next value in the following file is also between L1 and L2, then this value becomes the next F and another difference is saved. As soon as a value in the following file becomes greater than L2, then L1 becomes L2, and L2 becomes the next value in the lead root. This process is continued until the end of the file is reached. See the illustration below for a visual description of the difference:



d1, d2, d3, and d4 show the first four differences that will be saved from these two files.

Phases

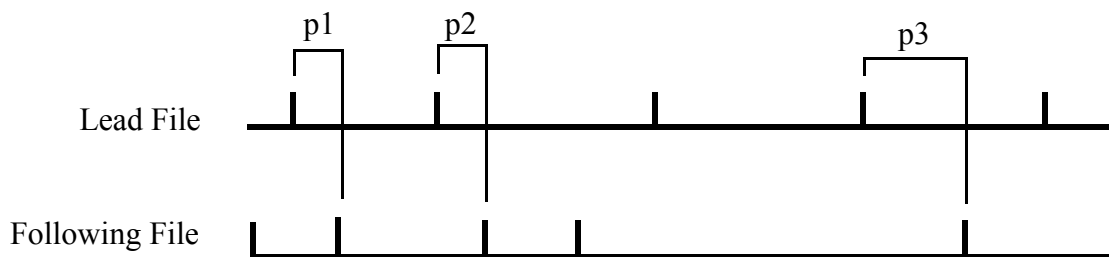
The phase is the difference divided by the period between the first and second data points on the lead file. They indicate how far displaced the data points in one file are relative to the data points in another file, as a percentage of one cycle of activity. If the phase between the data points in two files is fairly constant, this may indicate some coordination between the systems which generated the two data sets. There are several routines for phase calculation included in I Ching, and these are discussed separately below. You can choose which phase calculation routine you would like by selecting 'Phase Type' in the 'Options' menu (Section 6-4).

Original Phases

The first calculation routine is called Original Phases, and starts as was described above for differences. If the first and second data points in the lead file are given as L1 and L2, then the data points in the following file are searched until one is found that is between L1 and L2 (call this F). The first phase is then $(F - L1) / (L2 - L1)$. F is then set to the next data point in the following file. If F is still between L1 and L2, then another phase is saved as above. If F is more than L2 then L1 and L2 are both shifted by one data point (L1 becomes L2, L2 becomes the next data point), and the process is started over. See the diagram in the differences section (above) to see how phases would be calculated. The phases would be the differences shown, divided by the time between the data point on the lead file used for the difference calculation and the next data point on the lead file.

One-to-One Phases

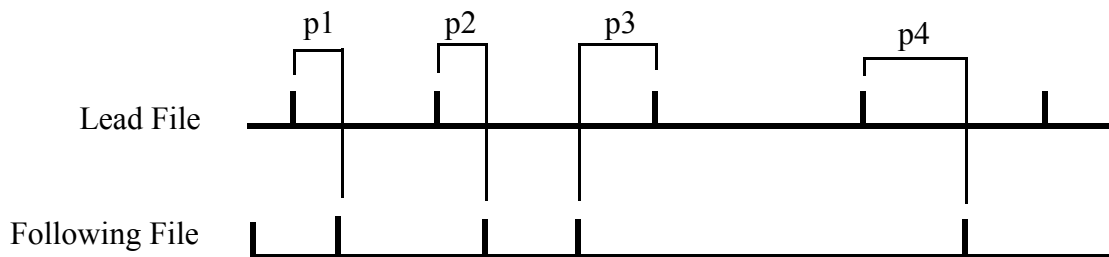
The calculation routine for One-to-One Phases is similar to that for original phases except that a maximum of one phase is found for each data point in the lead file. Thus, if there are two data points L1 and L2 in the lead file, and more than one data point in the following file lies between L1 and L2, a phase is calculated only for the first of the data points in the following file, and no phase is calculated for any other data points which lie between L1 and L2 in the following file. The diagram below illustrates this situation:



Notice that there is now no second phase calculated from the same lead file data pair that p2 is calculated from (Compare this diagram to the one in the differences section).

Lockstep Phases

Lockstep phases are a little more simple-minded than any of the other phases calculation algorithms. They should be useful when you see a one to one correspondence between events in the two files, but the phase varies around 0 or 1, so that other algorithms might not preserve the one-to-one correspondence. If the first data point in the lead file is L1, the lockstep algorithm finds the first data point on the following file which occurs after L1. Lets say this is F2, the second event in the following file. It then calculates the phase between F2 and L1, L2, according to the formula used above: $(F2 - L1) / (L2 - L1)$. Then the algorithm advances one data point in each file and calculates the phase for F3 and L2,L3, $[(F3 - L2) / (L3 - L2)]$, regardless of the position of F3 relative to L2 and L3. The lockstep algorithm keeps advancing through both data files one data point at a time calculating phases regardless of the positions of the data points relative to each other.



Notice that p3 will be negative.

This algorithm will perform very badly if occasionally one file has an extra or a skipped event relative to the other - even one extra or skipped event will completely throw it off. However, if you are careful about this, the lockstep algorithm may in some cases be the most meaningful phase calculation algorithm

Clean Phases

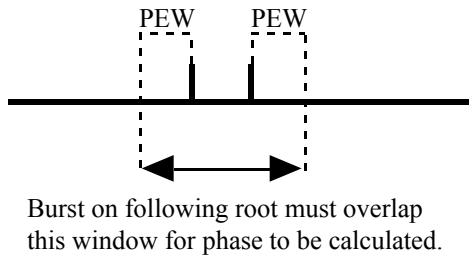
This is the most complex phase calculation algorithm of the four. Clean phases were designed to correct deficiencies in the other phase calculation algorithms when analyzing certain types of data. It is not clear that this algorithm is any better than the others, and it may have deficiencies of its own, but it is included here in case someone finds it useful.

Clean phases is designed to process data that is "clean", where the phases are approximately the same over the whole stretch of data, and where you do not want to include phases which are not approximately the same. It should not be used for data where the phases will vary a lot, or where there may be drift in the phases. The reason it is valuable to throw out phases which vary from the average in the phase calculation algorithm itself (as opposed to deleting them during post-processing) is that when a data point from the lead file is matched incorrectly with a data point from the following file during phase calculation, it may throw off the data points used in subsequent phase calculations. This algorithm is designed to avoid propagation of errors. Whether it actually achieves this goal is unclear.

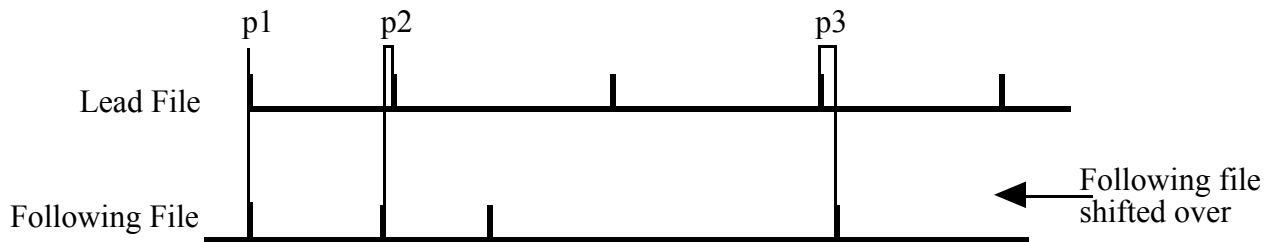
The first step in the clean phases algorithm uses the original phases algorithm (above) to find 5 phases in a row which are approximately the same. How approximately they must be equal is given by the Phase Match Slop variable, which is user adjustable. Each of the five phases must be within Phase Match Slop of the phase which precedes it.

Once these five phases have been found, the data points in the following file are shifted in time so that they line up with the data points in the lead file. In other words, they are shifted over by the phase found above, so that now the phases between the lead and following files should all equal approximately 0.

Once the data points are lined up, the algorithm goes back to the beginning of both the files, and finds pairs of data points which match up with each other. How closely they must match up with each other is given by the Phase Event Window, which is user adjustable. The onset from the lead file must be no more than the Phase Event Window (PEW) past the offset from the following file, and the offset from the lead file must be no more than the Phase Event Window in front of the onset from the following file. This establishes a window around the event from the lead file within which a piece of the event from the following file must fall. If the following event does occur in this window, a phase is calculated as above, and the length of the shift is added to the value of the phase. If not, then the next data point from either the lead or following file is examined and the comparisons are repeated. The diagram below shows the window, with the two thick vertical lines representing the onset and offset of a event on the lead root.



The diagram below illustrates the way that the algorithm should work - for clarity, only onsets of the events are shown:



Note that this diagram shows the arrangement after the shift of the following file. The shift would be added back into each of the phases calculated.

Histograms

When the 'Make Histograms' function is selected, a dialog box will appear asking for the parameters of the histograms. On the right you can decide what type of processing you would like to use to make a histogram. If you already have a processed file, and you would like to make a histogram out of the data in this file with no further processing, select Other Data. On the right you select the number of bins you would like, and the range of data values for these bins. Each bin will be a width of $(\text{End Bin} - \text{Start Bin}) / \text{Number of Bins}$. Further down are some check boxes which are used to further specify the histogram. If '< Min' is checked then an extra bin will be set up to hold any values which are less than the Start Bin, and similarly if '> Max' is checked then another bin will hold any values more than the End Bin. If wrap around is checked then any values which are exactly equal to End Bin will be added to the first bin instead of the last one. This is useful for phase histograms, since a phase of 1.0 is the same as a phase of 0.0. If 'Save Categories' is checked then when the histogram file is saved two columns will be saved, the first giving the range of the bin and the second giving the number of data values which fell in that bin. The '< Min' bin will always be put first and the '> Max' bin will always be put last when they appear.

Histograms are always saved to a file. In addition, histograms allow you to get chi square values for data other than phases. However there is currently no option to get the chi square without saving the histogram.

Filtering Files

This function is used to separate onsets from offsets in a file from Spike Studio. It will take an input file that has onset/offset data, and save an output file which has only the onset, only the offsets, or only the midpoint between each onset and offset. A dialog box will appear when this function is selected asking which of these three you wish to save. No statistics are found when filtering files.

Finding Statistics Without Processing

You may want to get statistics on a file of data without processing it in any of the ways discussed above. You can do this using the 'Get Stats' function. The 'Get Stats' function is only active when the Data Type is set to Processed Files (see Data Type, Section 6-2). When this is selected, all statistics will be found for the input files and printed on the screen (except for chi squares - see histogram section above). When you use 'Get Stats', no files are saved. Means, medians and standard deviations will be done using the standard formulas, not the formulas used for phases (see Mean, Median, or Standard Deviation).

Statistics

Whenever a data file is processed by I Ching, various statistics from the processed data are printed on the screen. Quite a few statistics are included in I Ching, and you can select which of these you would like to have appear and what their order should be. The statistics can be saved as a text file, and can be edited either in I Ching or in another program. Creating, saving, and editing these files is discussed in the Word Processing section. The statistics themselves are discussed here.

The statistics are displayed on the screen under a line of headings, one for each statistic. If you were to analyze a file for periods and then find phases for another couple files, you might get a screen looking like:

| Filename | Type | Mean | Std-Dev | Co-Var | Median | Q-1/4 | Q+1/4 | Chi-Sq | #-Datum | Cut |
|--------------------|------|-------|---------|--------|--------|-------|-------|--------|---------|-----|
| R06.04/19/05.4 | per | 1.423 | 0.059 | 0.041 | 1.350 | 1.286 | 1.395 | - | 227 | no |
| L15-R27.04/04/90.3 | pha | 0.307 | 0.285 | 0.927 | 0.297 | 0.003 | 0.637 | 19.677 | 606 | no |

Each of the statistics available is discussed below.

Selecting Statistics

To select which statistics will appear and what their order will be, use the 'Set Statistics' command in the 'Statistics' menu. A dialog box will appear with the names of all the available statistics and a box next to each one where you can type a number. The numbers indicate the order in which you want the statistics to appear. The statistic with a 1 in its box will appear first, followed by the statistic with a 2 in its box, and so on. Any statistic whose box you set to 0 will not appear. Be careful that you don't have duplicate or missing numbers or the program will complain.

Every time you set which statistics you want, this is automatically saved in the configuration (see Saving the Configuration), so when you open a new statistics window or the next time you use the program, it will default to showing the same statistics as you used the last time.

Filename

The filename given here is the filename under which the processed data is saved (assuming the data is being saved). If the data is not being saved, than this filename is the name of the input file. With phases or differences, the leading label and root are placed at the beginning of the name, followed by a hyphen and then the filename of the following root.

Type

The type specifies what type of processing is being done on the data.

| | |
|--------|----------------------------------------|
| 'per' | periods |
| 'pha' | original phases |
| 'opha' | one-to-one phases |
| 'lpha' | lockstep phase calculation |
| 'spha' | clean phase calculation |
| 'dif' | differences |
| 'len' | event lengths |
| 'unk' | anything else, including no processing |

Number of Points

The number of points gives the number of data points calculated from the processing of the input data. For instance, in calculating periods, if there were 250 onsets then the number of datum will equal 249, since there is one less period than there are onsets.

Mean

For all processing except for phases, this is a standard mean. For non-phase files, its:

$$\frac{1}{N} \sum_{i=1}^N X_i$$

For phases, the mean is calculated a little differently. Because phase values lie on a circle rather than on a straight line (1.0 = 0.0, and all values are between 1.0 and 0.0), just adding all the values and then taking a mean might give an erroneous result. For instance, if the phase between the two roots was hovering around 0.0, with some of the phases showing a slight lag (0.95 for instance) and some showing a slight lead (0.05 for instance), the mean should show that both roots are firing at about the same time, so the mean should equal 0.0. However, if the mean is calculated in the standard way, it will equal approximately 0.5, which is quite misleading. Therefore, the means, medians, and standard deviations for phases are all calculated on a circle using complex numbers.

This process begins by taking the sine and cosine of each phase, which turns the value of the phase into a complex number. Then the real and the complex parts of all the phases are averaged separately and converted back into a real number between 0 and 1. The actual algorithm in pseudo-code is given in Section 9-3.

Standard Deviation

The standard deviation is done in the standard way for all data except for phases. For non-phase files, the standard deviation is:

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N (\mathcal{E}_i - \bar{\mathcal{E}})^2}$$

As described for the mean, standard deviations for phases are done on a circle. For the standard deviation, this means that the phases are considered to be lying on a circle of circumference 1, and that the shorter distance around the circle is taken between each phase and the mean phase when calculating standard deviation. Thus the difference from the mean can never be more than 0.5. For example, if the mean phase is 0.8 and one of the phases is 0.1, then the difference between these two would be 0.3, not 0.7. The actual algorithm (in pseudo-code) is given in Section 9-3.

Standard Error

The standard error is done in the regular way for all data except phases. For non-phase files, the standard error is:

$$\sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (\mathcal{E}_i - \bar{\mathcal{E}})^2}$$

As described for the mean and standard deviation, standard errors for phases are done on a circle. For the standard error, this means that the phases are considered to be lying on a circle of circumference of 1, and that the shorter distance around the circle is taken between each phase and the mean phase when calculating standard error. Thus the difference from the mean can never be more than 0.5. For example, if the mean phase is 0.8 and one of the phases is 0.1, then the difference between these two would be 0.3, not 0.7. The actual algorithm is similar to the pseudo-code given for standard deviations given in Section 9-3.

Coefficient of Variation

The coefficient of variation is the standard deviation divided by the mean. Since this is not meaningful for phases, no coefficient of variation is given for phase calculations.

Median

The median is calculated by sorting all the data and finding the data value which is precisely in the middle.

As discussed under Mean and Standard Deviation, the median for phases is found by assuming the data values are on a circle. This is done as follows: instead of calculating the median by starting from the smallest value and counting up until halfway through the phases, the algorithm starts at the phase which is most nearly opposite of the mean. The algorithm finds the mean plus 0.5 (if this is more than 1.0 then it uses the mean minus 0.5), and then finds the next highest phase and starts counting from that phase. The actual algorithm is given in Section 9-3.

Quartiles (Q-1/4 and Q+1/4)

The quartiles are a way of showing variance around the median. Just as the median is the midpoint of the data set, the quartiles are the midpoints of the first half and the second half of the data set. Q-1/4 is the data value that is the midpoint of the first half of the sorted data set, and Q+1/4 is the midpoint of the second half of the sorted data set, where the halves are separated by the median. All else is the same as in the median calculation.

Chi Square

The chi square is a number which indicates whether a certain distribution of values is significantly different from a random distribution. It is calculated by putting all the data values into bins of equal width, assuming that a random distribution would place equal numbers of values into each bin, and summing the squares of the differences of the number of values in each bin minus the expected number of values in each bin for a random distribution. This sum is then divided by the expected number of values in each bin. The higher this figure, the farther away from random the distribution was. For an exact estimate of how far the distribution from random was, chi-square tables are available which have the percent chance of a particular chi-square value being due to a random distribution.

Since the chi square involves putting the numbers into bins, it is only done for phases and histograms. When phases are calculated, chi-squares are figured using 20 bins, each 0.05 seconds long (ie. the first bin is 0.0 - 0.05, the second bin is 0.05 - 0.1, etc.), and any 1.0 values are put in the 0.0-0.05 bin. When histograms are being calculated, the bins used for the chi square are the bins set up for the histogram. With all other types of processing, there is no natural range over which to set up bins, so no chi square is calculated.

One problem with the chi square test is that the significance of the value depends on the number of data points used to calculate the chi square value. Since different data files are likely to have different numbers of data points in them, you may wish to use the Cut function to standardize the number of data points used to determine the chi square values.

Sum

The sum is a sum of all the data values which were calculated from the processing of the input data.

Sum Squares

The sum squares is the sum of the squares of all the data values which were calculated from the processing of the input data. The formula is:

$$\sum_{i=1}^N X_i^2$$

Cut and Cut Side

If you have asked for the data to be cut, then Cut specifies how many points were cut before doing the statistics. Cut Side specifies whether points were cut from the beginning (specified by a 'b') or the end (specified by an 'e'). See Section 6-5 for a description of how to use the cut function. If no cutting was specified, then 'None' is printed under Cut.

Options

The 'Options' menu in I Ching contains many items which help control the processing. The 'Options' menu is where you select whether you want to process files by file, by condition, or by experiment, as well as selecting what part of the input files to use and whether to save an output file from the processing. You can also select between different phase calculation algorithms, and set up data cutting. The 'Set Roots for Phases' item is discussed elsewhere (Section 3-5).

Operational Unit

I Ching can process one file at a time, but its real power is its ability to process many files at once with a single command. Whether it processes one file at a time or many at once is determined by the Operational Unit which you select. The top three items in the 'Options' menu are the three different operational units which you can select, 'File', 'Condition', and 'Experiment'. If 'File' is selected, (indicated by a check mark), then only a single file will be processed each time you select one of the processing commands. On the other hand, if condition or experiment are selected then multiple files may be processed each time a processing command is given. When 'Condition' is selected, then the condition number is read from the input file that you select, and all files in the current folder with the same condition number and body as the input file will be processed. If 'Experiment' is selected then all files in the current folder with the same body as the input file you selected will be processed. Note that for processing by condition or experiment, all files must fit the filename template (Section 3-4). If a file does not have an appropriate name, you must process it individually by selecting 'File'.

Data Type

Files coming from 'Spike Studio' (the spike and event discrimination program for which this statistics package was designed) will generally be onset/offset files. All of the processing functions except Event Lengths only use one time for each event. You can select whether they will use the onset, offset, or the midpoint (halfway between the onset and offset) by using the 'Data Type' function. You can also elect to use each point in the file (e.g., if you have a midevent file, or a file of frequencies), or you can indicate that you are finding statistics for a processed file (see below).

When you select 'Data Type', a dialog box will ask which type you would like to use. If you select Onset, then every other number will be used starting with the first number in the file. If you select offset then every other number will be used starting with the second number in the file. If you select midpoint then the average of every pair of numbers in the file will be used. If you select Each Point then each point in the file will be used. Processed indicates that the data in the input file is already processed, and you just want statistics from it. For instance, if you Get Periods with Processed data, then the program will read the file you select as a file of periods and give you statistics on those periods.

For phase and difference calculations you have two additional options which are shown on the right side of the ‘Data type’ dialog box. You can calculate the phase of the offsets of the following root relative to the onsets of the lead root (‘Onset to Offset’) or you can calculate the phase of the onsets of the following root relative to the offsets of the lead root (‘Offset to Onset’). Selecting either of these does not change the data type which will be used for processing analyses other than phases and differences.

Save Options

Save options allow you to specify what types of processing you would like to have saved. Initially no files will be saved from the processing. To save the data from the processing, select the 'Save Options' item in the 'Options' menu. A dialog box will ask which types of processing you would like to save. Check the processing types you want to save, and uncheck those you don't. Whenever a processing command is given to the program, I Ching will check whether you have requested for that type of processing to be saved, and if you have then it prompts you for information about where to save the file.

If the 'Time Stamp' box is checked then two columns of numbers will be saved in each output file. The first column will contain a time stamp for each data point in the second column. A time stamp is generally the onset of the first event involved in the calculation of that data point. For instance, when processing for periods, the time stamp would be the onset of the first event used in calculating that period. Time stamps can be saved with any of the types of data shown on the left of the dialog box, and may be particularly useful for graphing how certain characteristic of the data changes over time.

Phase Type

As explained above under phases (Section 4-4), there are four phase calculation algorithms currently implemented in the program, each with its own strengths and weaknesses. You can select between these using the 'Phase Type' command. When you select this command, a dialog box will ask which algorithm you would like to use. Whichever algorithm you select will be the one that is used whenever you process for phases.

Cutting

You may want to use a fixed number of data points in your calculations, or you may sometimes have a situation where the beginning or the end of your data file is not very good. In either of these case you may wish to trim numbers off either the beginning or end of your data file. To do this, use the 'Cut Data' function in the Options menu. A dialog box will ask how many data values you would like to use, and whether to cut off extra values from the beginning or the end of the file. From then on, whenever data is processed, if there are more processed values than were specified in the Cut dialog box then the extra values will be cut off whichever end you specified.

Cutting will only affect the processed data as it is used for statistics. If you are saving the processing, then all the processed data will be saved to the output file, with no cutting performed on the data in the output file. However, if histograms are saved then they are saved with only the cut data placed in the bins. In any case, if there are less processed data items than were specified in the Cut dialog box, all the data are used.

Saving the Options

You can save the settings in the program so that you don't have to reset them each time you run the program. The settings are stored in a place called the configuration, and to save them you use the 'Save Configuration' command in the 'File' menu.

When you run the program all the settings in the options menu will be the same as they were when you last did a 'Save Configuration'. Note that the order in which statistics appear is automatically saved to the configuration every time you do a 'Set Statistics' (Section 5-1).

Word Processor Functions

The core of I Ching is a simple word-processor, and includes many of the standard word processor functions. You can type in text, cut, paste and move text around, open text files, and change fonts, styles, and type sizes. Most of the functions in the 'File' menu, all of the 'Edit' menu, and the 'Font', 'Size', and 'Style' menus are equivalent to these menus in other Macintosh programs. For example, to type in new text, (a comment for instance), move the cursor to the place that you want to add the text and start typing. Below are brief descriptions of a few of these functions, but since they are very similar to standard Macintosh commands I am not going to explain them extensively.

New Statistics File

When the program is started, a window will automatically be opened where statistics are displayed. If, during the course of a session with I Ching, you would like a fresh window for statistics, select the 'New' command from the 'File' menu. This will open up a new window on top of any window that was already open, and any new statistics will be written into the new window.

Saving Statistics Files

To save the text in one of the statistics windows, make that window active (by clicking in it) and then select 'Save' or 'Save As...' from the 'File' menu. A standard Macintosh save requester will ask what file name to save the statistics under. This file will be a text file, which can be read into any program that can read text files (e.g., any word-processor, stats or graphing programs, spreadsheets, etc.).

Opening Files

You can open and view any text file using I Ching. Simply select the 'Open' command from the 'File' menu, and then select the file you wish to open. You can look through the file, edit it, save any changes, even use it as a place to put statistics. (The statistics are always added at the end of the currently active window). One note of caution - there is a limit of approximately 30000 characters allowed in each file that the program opens, and the program will slow down quite a bit well before that, so generally only small text files should be opened by I Ching.

The file opening feature is very useful for looking at the data files. Both the onset/offset files and the processed data files can be opened and examined using I Ching.

Why is the Font Courier 10 Included on the Disk?

The preferred font for I Ching is 10 point Courier because the characters are all the same size, so it looks much more orderly than other fonts. If you don't have that font on your Mac, you can install it from the 'Courier 10' file on your I Ching disk.

Help

The help function repeats a lot of information from this manual. You can use it to get help for a command without leaving the program. *For the help command to work, a file called 'I_Ching_Help' must be in the same folder as I Ching.* Select 'Help' from the 'File' menu and a window will be brought up that is divided down the middle. On the left are the different subjects about which you can request help. On the right is the help. When you select one of the items on the left, the text on the right will change, and a new list of items may also appear on the left. These new help items are sub-categories of you're original selection. To get back to the original selection, select 'Previous Help'. To leave help, click in the close box of the help window.

In general, the manual will be more thorough than the on-screen help and hopefully better written, so if the help doesn't answer your question, read the manual. The help feature will reference the appropriate part of the manual. It is possible to modify the text in the help menu. The text is stored in the I_Ching_Help file mentioned above. This is a text file which can be edited with a word processor (including I Ching itself). The format which help items must take is as follows:

```
§Main Menu Items§
Subitem 1
Subitem 2
Subitem 3
```

This is where you put the text which should appear when Main Menu Item is selected.

```
§§§§
§Subitem 1§
sub-subitem a
sub-subitem b
```


Text for Subitem 1.

§§§§

The title of each item about which help can be requested is put between two §'s (typed by holding down the option key and pressing 6). Below the title are all the related items which the user might want to access. A blank line follows, and then the text that should be presented for the item, if any. Finally comes four §'s to end that help item.

Programmers Appendix

As mentioned in the introduction, source code is available for those who would like it. In general, it will be faster and easier to have the author modify the program for you, but you are welcome to try it yourself. This section gives a very brief introduction to the layout of the code. It also presents pseudo-code for the statistics using complex numbers.

I Ching Squared is written in C using Think C's Object Library to implement its user-interface. I am not going to discuss this library at all, and you will have to have Think C version 5.0 with the library to do any programming of I Ching (and also know C, of course). Knowing how its object library works would be a big advantage, but might not be necessary (and the learning curve is steep, as I found out the hard way). The code itself is well-organized, and commented alright (by my standards, anyway).

The code is divided into several parts. The commands in the 'Statistics' menu are handled by commands in the file `C_IChingPane.c`. All other commands are handled in the `C_IChingApp.c` file. The processing itself is done by functions in the `I Ching Funcs.c` file, as are all the calculations of statistics. This is also where the input-output functions for data files are. A separate set of files has commands to handle the 'Help' command, another handles the 'About I Ching' command, and the `C_IChingDoc.c` file takes care of opening and closing file and keeping track of windows. Where some of the other stuff like printing happens is a mystery to me. It works, so I've never bothered to find out where it works.

There are probably two things you might want to modify in the program. One is to add in another type of data processing, and the other is to add in a new statistic. Both of these are outlined below. Good luck (and unfortunately, you'll probably need it).

Adding a Processing Command

The first thing you should do to add a processing command is to modify the I Ching menus using ResEdit to add in a menu item for your new command. Add it into the Statistics menu. You will notice that all the menus have a number sign and then a number after the name of the menu item. This is the number of the command which will be given to the program. Pick an appropriate number (look at the numbers of other menu items). You must then go into the `C_IChingApp.h` file and add in a `#define` statement defining the number of the new menu item. You will also need to add in another `#define` statement here which defines a number identifying this new type of processing. You will see a collection of `#define` statements giving numbers for `PERIODS`, `PHASES`, `BURST_LENGTHS`, etc.. Notice that the numbers are all multiples of 2 (and make your's one also).

There is a standard set of functions which handles all processing commands in a uniform way. The command is first handled by the `DoCommand` function in the `C_IChingPane.c` file. This function finds the input and output file names, and runs through all the files which should be processed. You must add the command you just created to the case statement at the top of this function. If there are any special information you need to get before processing (such as in the `Make Histogram` or `Filter Files` commands), add in an `if` statement near the beginning of the function which brings up a dialog box to get the necessary information.

With a little luck, you should not need to modify this function any more, and it will correctly ask for input and output file names, and run through all the appropriate files. The next function called is the `GetStats` function, in the same file, which is called for each file. You will probably not need to modify this function at all. It's main purpose is to call the `Read_Data` function.

The `Read_Data` function is where the actual processing occurs. It is contained in the `I Ching Funcs.c` file. This is where you insert your processing code. There is a big `if/else if` statement which includes all the different types of processing, so insert an `else if` statement for your type of processing. All processing is done locally in `I Ching`, meaning that you can compare data values to the ones just in front or behind them, but you cannot compare widely separated data values in your processing without a lot more added code. To see exactly how to write the processing code, what input/output functions to use, and where to put the results, look at the code for one of the other types of processing (periods and event lengths are good examples).

If you want to save data from this processing, you will need to go to the top of the `Read_Data` function and add in the name of your new type of processing to an `if` statement there. You might also want to add to the `if` statement in the `ComposeName` function.

Finally the statistics are calculated by the `GetStats` function, and the results are passed to the `PrintInfo` function (just below `GetStats`). Again, there are a couple `if` statements in `PrintInfo` which you may wish to modify.

Adding a New Statistic

The statistics functions are all in the I Ching Funcs.c file. You can look at any of the others to see how to make one. The other places you will need to add code are in the GetStats function and the PrintInfo function. The GetStats function calls the different statistics functions, so you will need to add in a call to your new statistic. The results from the statistics are stored in a data structure called 'info', which is defined in C_IChingApp.h, so you should add in an item to this structure to store your statistic. The final thing you will need to do is to add in some code to the PrintInfo function to print your statistic. One slight problem will be to find space for the statistic. The statistics currently take up the width of a standard page, so unless you are not concerned about printing the statistics you may have to remove one of the old statistics and replace it with yours. Another option would be to print sideways.

To change the statistic titles at the top of the window, go to the I_IChingPane function in the C_IChingPane.c file. To expand the margins in the window, you're on your own (but it shouldn't be hard to figure out if you understand Think C's object library - heh heh heh...

Selected Pseudo-Code

This section presents pseudo-code for the statistics functions which find mean, standard deviation, and median for phase data. The reason these might be interesting is that they put the phase data on a circle before doing their calculations, so they are somewhat different than the regular mean, deviation, and median calculations.

Phase Mean

```

for(i = 0; i < num_datum; i++) {
    tot_x = tot_x + cos(6.283185 * data[i]);
    tot_i = tot_i + sin(6.283185 * data[i]);
}

length = sqrt(tot_x * tot_x + tot_i * tot_i);
if(length == 0) then return an error since there is now more than one possible avg
tot_x = tot_x / length;
tot_i = tot_i / length;
avg = atan2(tot_i, tot_x) / 6.283185; This atan gives a number between -pi and +pi, which we
                                     then divide by pi to get a number between -0.5 and 0.5
if(avg < 0.0) avg = avg + 1.0;         This converts the avg to a number between 0.0 and 1.0
return(avg);

```

Phase Standard Deviation

```

for(i = 0; i < num_datum; i++) {
    if(data[i] > mean + 0.5) tot = tot + Square(mean - data[i] + 1.0);
    else if (data[i] < mean - 0.5) tot = tot + Square(data[i] - mean + 1.0);
    else tot = tot + Square(data[i] - mean);
}
std_dev = sqrt(tot / num_datum);

```

Phase Median

```
shift = mean + 0.5;
if(shift > 1.0) shift = mean - 0.5;
sort(data);  Sorts the data so it goes from lowest to highest
ptr = 0;
while(shift > data[ptr] and ptr < num_datum - 1) ptr++;  Find the first data point above the shift
for(i = 0; i < num_datum / 2; i++) {
    ptr++;
    if(ptr == num_datum) ptr = 0;  Wrap around 1.0 to 0.0
}
return(data[ptr]);
```