R2 User's Guide

Version 1.1

James H. Steiger
Rachel T. Fouladi

University of British Columbia

R2 is a "legacy" computer program, originally writeen for MSDOS computers in 1992, which implements some statistical procedures not generally available for inference on multiple correlation coefficients. R2 allows:

- Calculation of exact confidence intervals, probability levels, and lower confidence bounds for the squared multiple correlation.

- Power calculation for tests of significance of the squared multiple correlation.

- Calculation of the sample size necessary to achieve a desired level of power for testing a hypothesis of zero multiple correlation.

- Testing hypotheses of the form $P^2 = a$, for values of a other than zero.

- Direct tabulation of percentage points of the distribution of R^2.

- Calculating the exact cumulative distribution function of R^2.

Multiple correlation coefficients are computed by all major statistical packages. Yet standard statistical software packages, such as SYSTAT, SPSS, and BMDP, do not offer a full range of inferential procedures in connection with the multiple correlation. Most notable by its absence is a confidence interval on the population coefficient of determination.

The cumulative distribution function for R^2 under the assumption that k variables have been sampled independently and randomly from the multivariate normal distribution is difficult (and was, in the era of mainframe computing, very expensive) to compute reliably. Exact confidence intervals for $\square$, though in principle simple enough to obtain, require several iterations, each in turn requiring calculation of the c.d.f. of R^2. Moreover, the numerical routines must also be constructed very carefully to remain stable under a variety of parameter choices. Perhaps this explains why the procedures implemented in R2 have not achieved popularity at this time.

R2 is provided as EXE files. R2 will run in a DOS window on versions of Microsoft Windows up to and including Windows XP. **Instructions for running R2 under Windows 7 are provided in a separate document**. R2 uses a math coprocessor if one is present, but a coprocessor is not required (although it is definitely recommended). The program is not copy protected.

At the heart of the program are two routines which calculate the cumulative distribution function of R^2. One (the routine "emphasizing accuracy"), based on theory described in Lee (1972, p.178), is accurate to at least 5 decimal places, the other (the routine "emphasizing speed") based on Lee's (1971, p. 123) non-central F approximation, is accurate to about 4 decimal places, and is satisfactory for most applications. The difference in accuracy between the methods is, in a practical sense, minor, but the difference in speed is substantial for some problems.

R2 has a simple yet sophisticated user interface, and runs on any MSDOS computer. It is completely menu-driven. After a procedure is selected from a menu, the user is asked to provide a series of values. The user types in these values. The requested statistical quantities are then calculated.

Installing R2

R2 as a zip file. The following files should be provided in addition to the documentation:

| File	Purpose
R2.EXE R2FOR.EXE R2BAS.MON	Program files.
GENERAL.HLP ALGO.HLP FILES.HLP OPTIONS.HLP	Help files. These are standard ASCII text files which can be edited and modified by the user.

The files R2.EXE, R2FOR.EXE, OPTIONS.HLP, FILES.HLP, ALGO.HLP, GENERAL.HLP, and R2BAS.MON must all be in the current DOS path when you are running R2.

A Tutorial

The operation of R2 is easiest to discuss in the context of a simple example. In what follows we assume you have copied the 7 program and help files into a directory on your hard disk, or you have made a non write-protected floppy drive with the R2 files your active disk drive.

1. Starting R2

If you are in Windows, open a DOS box. Be sure you are in the directory where the R2 files are located. Type "R2" followed by a carriage return to activate the program. At this point, you should see a title screen as follows:
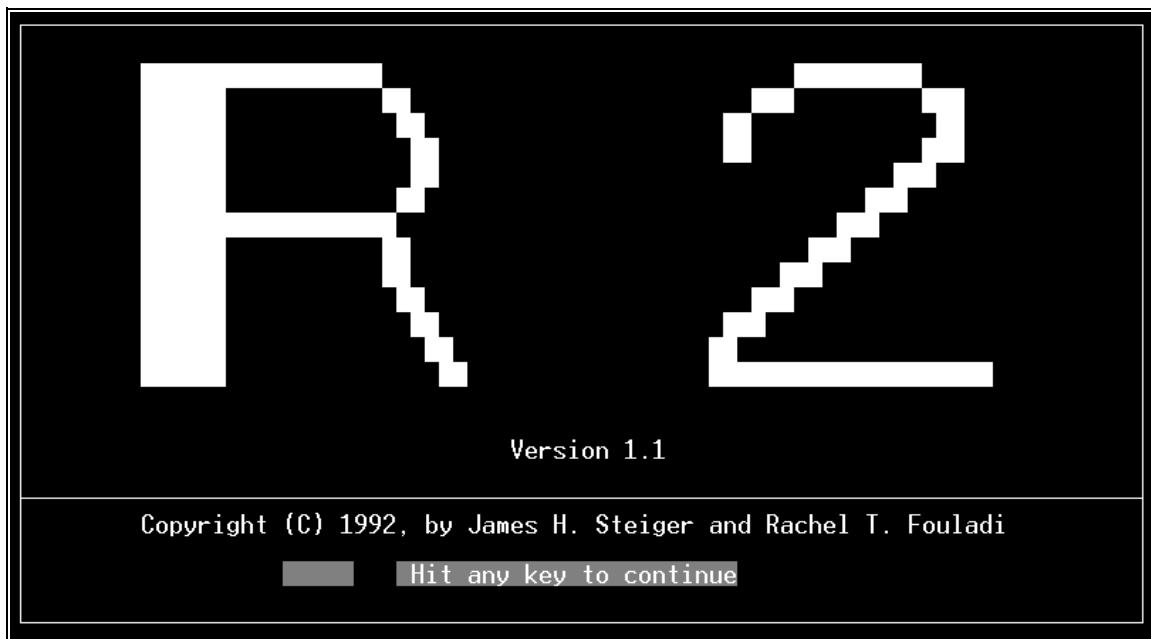
Hitting any key will activate the main program menu. For example, if you simply tap the space bar on your computer, you should see the following:
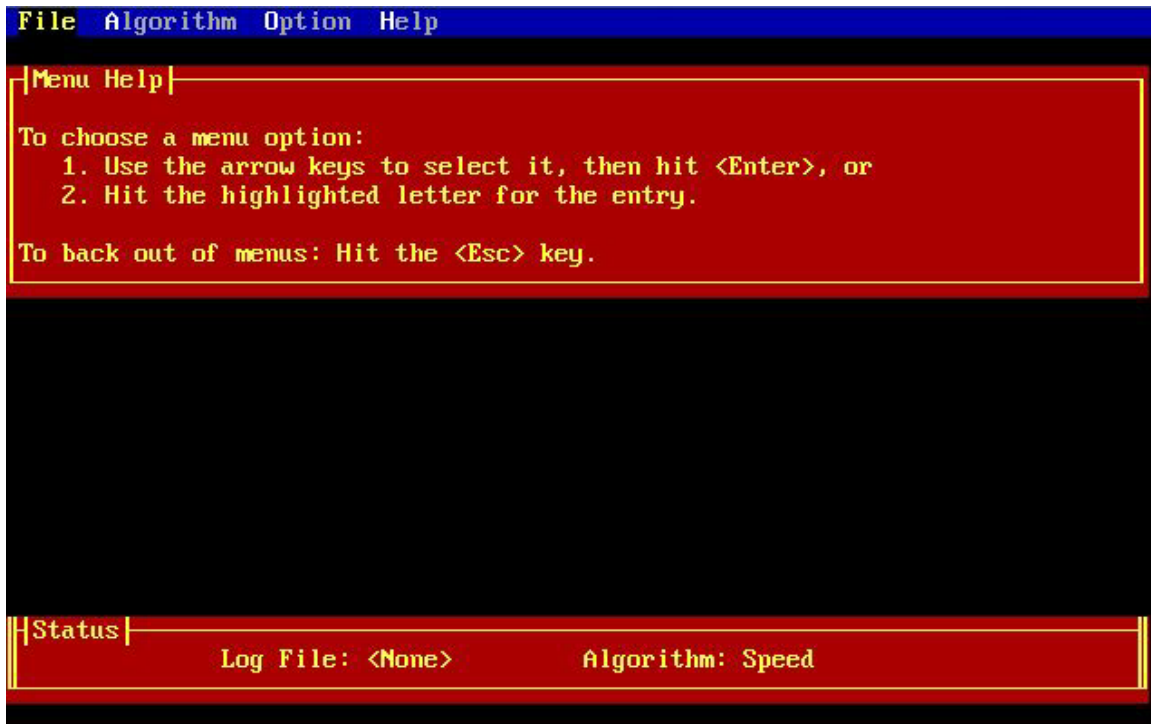
You are now ready to begin using the program.

2.   The R2 Menu and Help System.

At the top of the scren is a menu bar with the four main menu subdivisions, "File", "Algorithm," "Option", and "Help." At the bottom of the screen is a small area called the "Status Bar." We will discuss each of these areas in turn.

In order to operate R2, you have to know how to use its menu control system.

When the program first starts, a brief message titled "Menu Help" appears just below the menu bar, giving very brief instructions on how to use the menu system. (This message will disappear permanently after you activate your first menu choice.) The menu bar has four main menu subdivisions. The subdivision currently selected will be indicated by a reverse video highlight. When the program first starts, as in the figure above, the "File" option is selected. Each subdivision on the menu bar can be selected in two ways, as described in the Menu Help message.

Suppose, for example, you wished to activate the "Help" menu selection. You could press the right arrow button on your keypad 3 times to move the highlight bar over to the "Help" option, then hit the <Enter> key. At that point, a secondary "drop-down" menu would appear below the help selection. Another way of accomplishing the same thing is through the use of *key letters*. Each menu option has a single letter highlighted. If you

simply press that letter, you will activate that submenu. So simply pressing the letter "H" on your keyboard would also activate the "Help" submenu.

Generally, using *key letters* for an option works faster than using the arrow keys and the <Enter> key.   After activating the "Help" submenu, you should see the following:
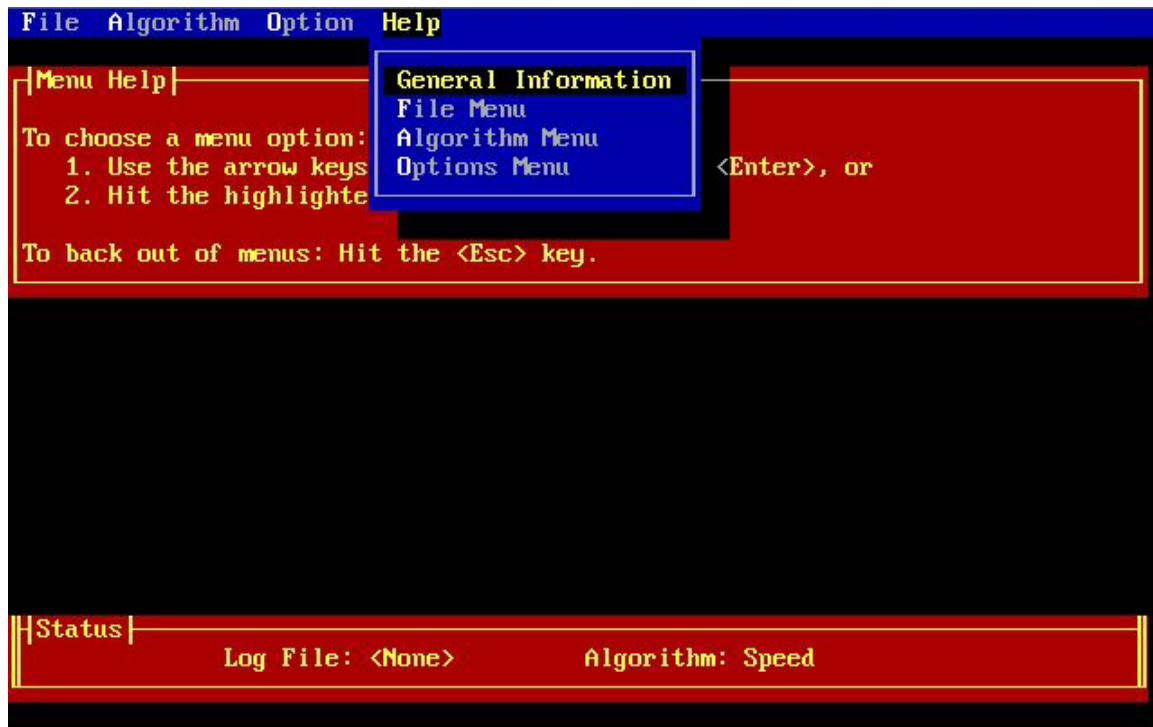
Once you have activated a menu subdivision, you navigate through its alternatives in much the same way you go through the main menu alternatives. To see how this works, press the down arrow four times. You will see the highlighted bar move down through the various entries. If you press the down arrow key after you get to the last selection, the highlighted bar will reappear at the top of the list.

To "back out" of a menu subdivision, hit the <Esc> key. Try it for yourself. Then hit the "H" key to reactivate the Help section. When you are in a submenu, you can move directly to other submenus by hitting the left and right arrow keys. To see how this works activate the "Help" drop-down menu, then hit the left arrow key 4 times. Notice how you activate each submenu in turn, and end up back where you started.

To continue the tutorial, move the highlighted bar to the "General Information" slot. Note that you can do this *either by using the up and down arrow keys, or by hitting the key letter "G" for that slot.*

At that point, you will enter the Help System, which will display general instructions for the R2 menu system. The screen should look like this:

```
File  Algorithm  Option  Help
  ┌─How to use Help├──────────────────────────────────────────────
        Use arrow keys, <PgUp>, <PgDown> to page through document.
        <Home>, <End> to go to beginning or end. <Esc> to quit.

                              R2
                          Version 1.1

                      Using the Menu System


      R2 has a menu system much like other programs with "drop-down"
  menus, like LOTUS 1-2-3 and Microsoft Windows applications. However, a
  mouse is not currently supported. At the top of the R2 screen are the
  main menu options. Their use is described below. At the bottom of the
  screen is a status bar. The bar shows whether a log file is currently
  in use, and which computing algorithm is currently being employed.
  While extended computations are in progress, a "COMPUTING..." sign
  will flash on and off at the lower left of the status bar.

 ║┤Status├─────────────────────────────────────────────────────────
          Log File: <None>            Algorithm: Speed
```

Brief instructions on how to page through the help messages are displayed above the file display area with the heading "How to use Help."

The Help system simply selects a specific help file from the current disk directory. These files have the extension .HLP and have names corresponding to the menu entries. For example, the file GENERAL.HLP contains the help information called up by hitting the "General Information" section.

These files are in a standard ASCII (text) format, and as such can be modified by most word processors and text editors. If you wish to add information to the help files which you find useful, please feel free to modify them in any way which suits you. However, if you choose to edit these files with a word processor, remember to save them as *plain text files*, otherwise they may cause problems when you try to read them. ***We strongly suggest that you not modify the Help files unless you have at least one backup copy.***

Try reading one or more of the help files, using the instructions under "How to use Help." Then hit the <Esc> key twice to reinitialize the menu system.

3.   The R2 Log File System

R2 offers you the option of saving a record of all the calculations it performs. The records are maintained in text files called "Log Files."

In many situations, you will not need a log file, because you will be using R2 to perform only a single calculation. However, if you wish to maintain a permanent, machine-readable record of the calculations performed by R2, you can use the File menu option and direct the program to maintain a log file. The file is an ASCII (text) file containing a record of the current session. This file can be typed to the screen, read through an editor, or printed. To open a log file, activate the "Open Log File" option of the "Files" menu. You will then be prompted for the file name of your log file. The file name should be entered *with no extension and no path.* The program will automatically add the extension .LOG to whatever name you provide.

Once a log file is opened, it remains open until it is closed or the program is terminated. R2 should never overwrite information already in a log file. If you open a log file which already exists, R2 will add any new results to the end of the information already in the file.

To practice opening a log file, select the "Open Log File" option and open a log file called "RESULTS." If you open the file properly, you should now see its name displayed at the bottom of the screen in the Status Area.

The status bar at the bottom of the screen should now state "Log File: Results."

4. Choosing an algorithm.

Calculations performed by R2 require use of a routine for computation of the cumulative probability distribution function of the sample squared multiple correlation. This is a difficult calculation to perform both quickly and accurately. R2 offers two very accurate procedures. One emphasizes speed while sacrificing a small degree of accuracy. The other gains accuracy while sacrificing a considerable amount of speed. For almost all calculations of practical interest, the routine which emphasizes speed will be sufficiently accurate that it will agree to the 4th or 5th decimal place with the more accurate routine. Consequently, the faster algorithm is the default choice in R2. For several functions calculated by the program, the fast approximation is calculated first to provide a starting value for the slower, more precise method. In what follows we will assume that the faster method has been selected. When you activate the Algorithm menu option, the drop down menu appears with the currently selected algorithm indicated with an asterisk. Select an algorithm using the standard menu selection method. Whichever method you select will be shown in the status bar area. Press "A" to activate the Algorithm menu, and practice changing the algorithm back and forth. For the completion of the tutorial, leave the algorithm in the default mode (emphasizing speed).

5. A Sample Calculation: Exact Confidence Intervals and Confidence Bounds for the Squared Multiple Correlation,

A unique feature of R2 is its ability to calculate exact confidence intervals for a squared multiple correlation. This confidence interval is much more informative than a shrunken estimator, because it gives an indication of precision of estimation.

Confidence intervals for P^2 are difficult to compute. The interval, formally defined as two values, x_{low} and x_{high}, such that, over repeated samples,

$$\Pr\left(x_{low} \leq P^2 \leq x_{high} \right) = 1 - \alpha$$

is computed using the "inversion" procedure discussed by, among others, Cox and Hinckley (1974, p. 213). (Note: After this program was published, Steiger and Fouladi, 1997, published an extensive account of such confidence intervals that may be downloaded from http://www.statpower.net.) Basically, the procedure finds the endpoints of a $1-\alpha$ confidence interval for P^2 by finding iteratively those values of P^2 which place the observed value of R^2 at the $\alpha/2$ and $1-\alpha/2$ percentage points. R2 uses a fast secant-method root-finding routine to solve for the endpoints of the interval.

As an obvious specialization, one can also calculate a "statistical lower bound" for P^2. This is a value $x_{lower\ bound}$ such that
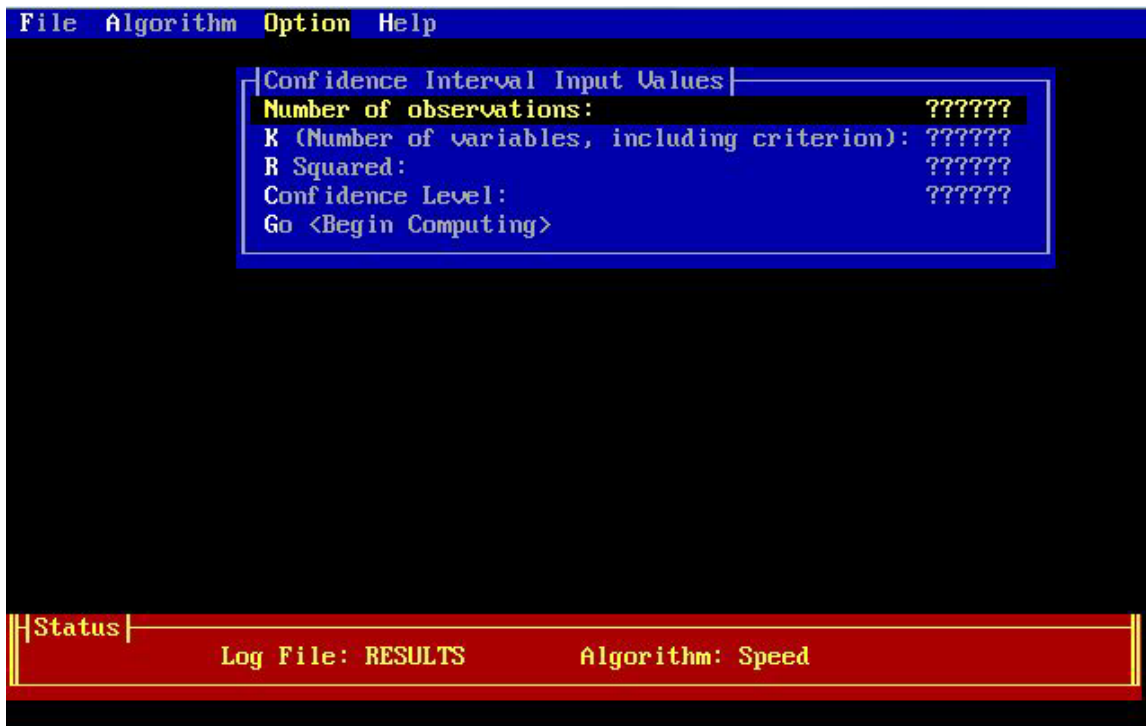
$$\Pr\left(x_{lower\ bound} \leq P^2 \right) = 1 - \alpha$$

Lee (1972) suggested calcuating such intervals from his tables, using interpolation. This is extremely cumbersome and somewhat inaccurate.
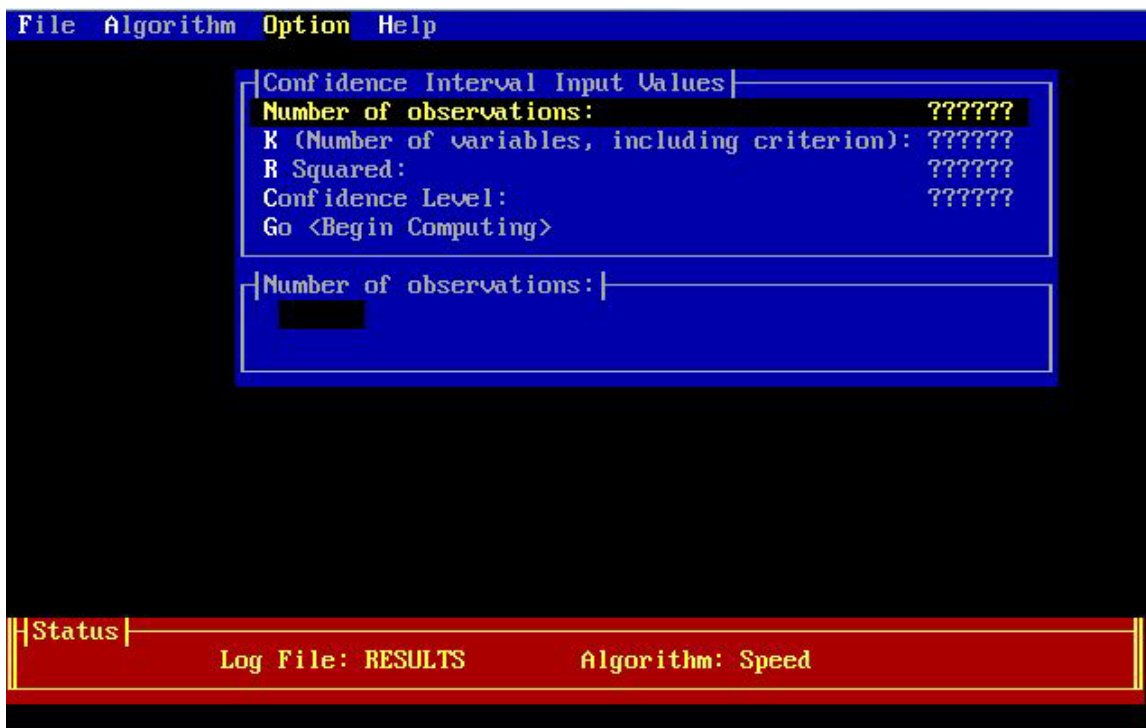
The time taken to calculate confidence intervals is an increasing function of N and R^2. Use of the faster percentage point routine is recommended if $N > 300$ or $R^2 > 0.8$. The program only accepts confidence values between .6 and .999.

Confidence intervals on the coefficient of determination can be quite revealing. Suppose, for example, 15 predictors are used, and $N = 104$, and a value of .50 is observed for R^2. Let's work through the calculation of a 95% confidence interval for this situation. Begin by selecting the "Options" menu and activating the drop-down menu. Then select the "Confidence Interval" option from the drop down menu and (if necessary) press <Enter>. (If you activate the Confidence Interval option by hitting the letter "C" you will not have to use the <Enter> key.)

If you have executed these instructions properly, a data entry menu will drop down and the screen will appear as follows:

```
 File  Algorithm  Option  Help
                ┌┤Confidence Interval Input Values├──────────────┐
                │ Number of observations:                  ??????  │
                │ K (Number of variables, including criterion): ??????│
                │ R Squared:                               ??????  │
                │ Confidence Level:                        ??????  │
                │ Go <Begin Computing>                             │
                └──────────────────────────────────────────────────┘

 ┤Status├
            Log File: RESULTS           Algorithm: Speed
```

You must now enter the requested values. To activate the data entry mechanism for any value, either enter the key letter for the value, or move the highlight bar to that value, then hit enter. Begin by pressing <Enter> (or the key letter "N") to activate the data entry for N, the number of observations. A data entry window will appear just below the drop-down menu, and prompt you for the desired value.

```
 File  Algorithm  Option  Help

          |Confidence Interval Input Values|
          Number of observations:                      ??????
          K (Number of variables, including criterion): ??????
          R Squared:                                   ??????
          Confidence Level:                            ??????
          Go <Begin Computing>

          |Number of observations:|

              ▮


 |Status|
              Log File: RESULTS        Algorithm: Speed
```

At this point you will be in *data entry mode*. In this mode, the following editing functions are available.

Editing Function	Key(s)
Cursor left one character	Left Arrow
Cursor right one character	Right Arrow
Cursor to beginning of line	<Home>
Cursor to end of line	<End>
Delete character at cursor	<Del>
Delete character to the left of cursor	<Backspace>
Delete entire entry	<Ctrl Y>
Toggle Insert Mode[1]	<Ins>
Abort Entry	<Esc>
Accept Entry	<Enter>

Type in the value 104. then press <Enter>. You will note that the value 104 has replaced the row of question marks. The value 104 will remain in this position until you replace it with a different value, or exit the program.

---

[1]"Insert mode" is the normal mode of operation. In this mode characters to the right of the character you type are moved over to the right. In this mode, the cursor is a solid block. You can toggle out of "Insert mode" into "overtype" mode, in which characters you type overwrite those already present. In this mode, the cursor will be an underscore character, i.e., "_".

Next enter the number of variables, *K*. Since, in our example, there are 15 predictors and 1 criterion, enter the value 16.

Enter a value of .50 for R^2.

Then enter a confidence value of .95.

At this point, you can begin the computation by pressing the letter "G" for "Go," or by moving the highlight bar to the "Go <Begin Computing>" row and then pressing the <Enter> key.

The message "COMPUTING..." will flash in the left corner of the status bar as shown below while the confidence intervals are computed iteratively.
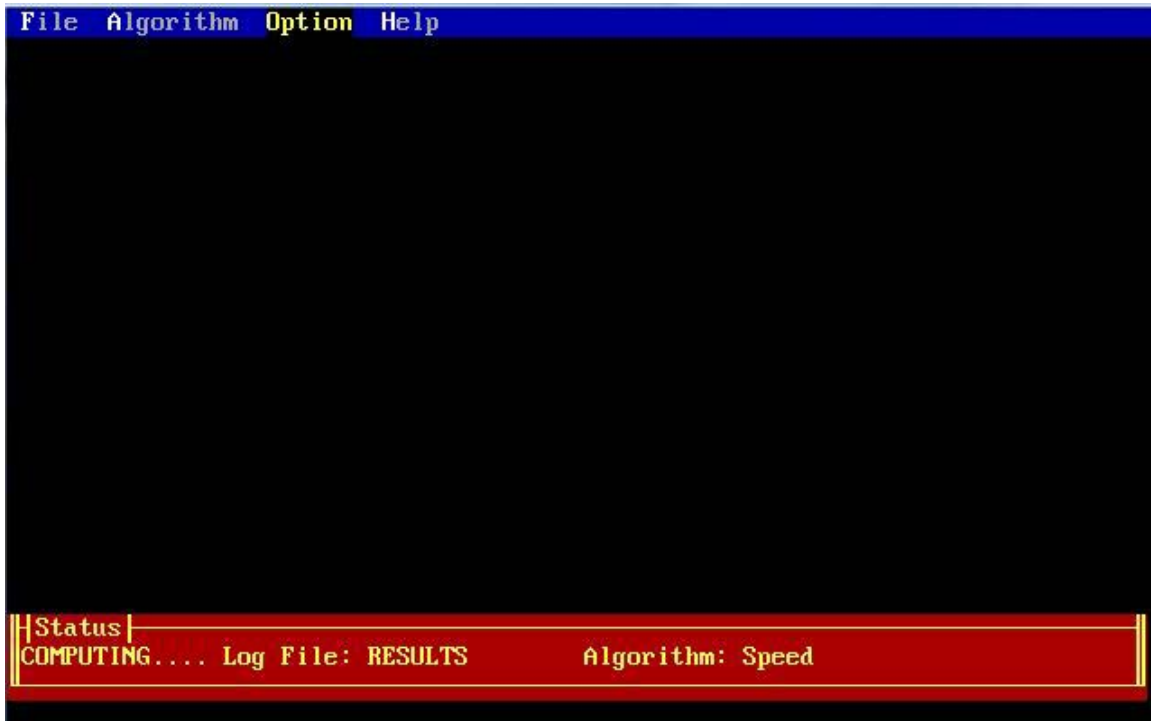
When the calculations are completed, the program will print the message "<Hit any key to continue>." Pressing any key at that point will cause the results of computing to be displayed.

In the current example, you will see the following:

In this case, we find that the sample size is insufficient to determine the population coefficient with a high degree of confidence. A 95% confidence interval ranges from .251 to .565. A 95% lower bound for the squared multiple correlation is .2785. The program also reports the "significance" level, or "Plevel" of the observed multiple correlation.

At this point you can record the data by hand. Remember that the computations are also being recorded in a file called "RESULTS.LOG." Press the <Enter> key to resume program operation.

To complete this brief tutorial, activate the "File" option from the menu and press "X" for "EXIT." This will close down R2 and take you back to the DOS prompt. At that point, you may wish to examine the contents of the log file for our session. Enter the command

TYPE RESULTS.LOG

and you will see something like the following.

```
----------------------------------------------------------------------

Output from R2 -- Version 1.1

Date: 02-03-1992
Time: 23:42:36

High Speed Algorithm is in use.
----------------------------------------------------------------------

Confidence Interval Calculation

N:                104
K:                 16
RSquare:        .50000
Confidence:     .95000
Lower Limit: 0.25135
Upper Limit: 0.56532
Lower Bound: 0.27852
Plevel:      0.00000
----------------------------------------------------------------------
```

This completes our introductory tutorial tour of R2. Hopefully, you have found the program fast and easy to use. In the section which follows we will describe some of the additional capabilities of R2.

Additional Capabilities of R2

1. **Percentage Points of the Distribution of R^2 .**

R2 can calculate percentage points of the distribution of R^2, eliminating the need for complex tables requiring interpolation. Suppose, for example, you wished to know the 95th percentile of the distribution of R^2 when $P^2 = 0.5$, there are 6 predictors, and sample size (N) is 125. Activate the "Percentage Point Calculation" option from the "Options" menu and enter the data. When you are done, the screen should look like this:
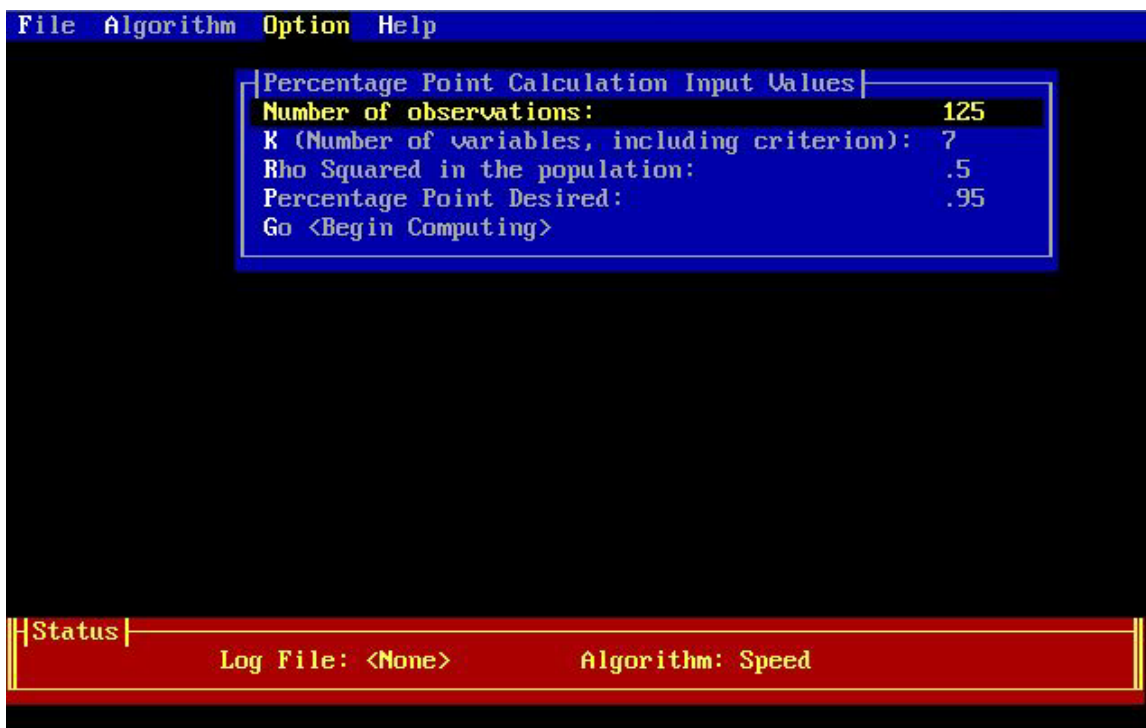
Press "G" to begin computing. You will find that, under these conditions, the 95th percentile is a squared multiple correlation of about .619.
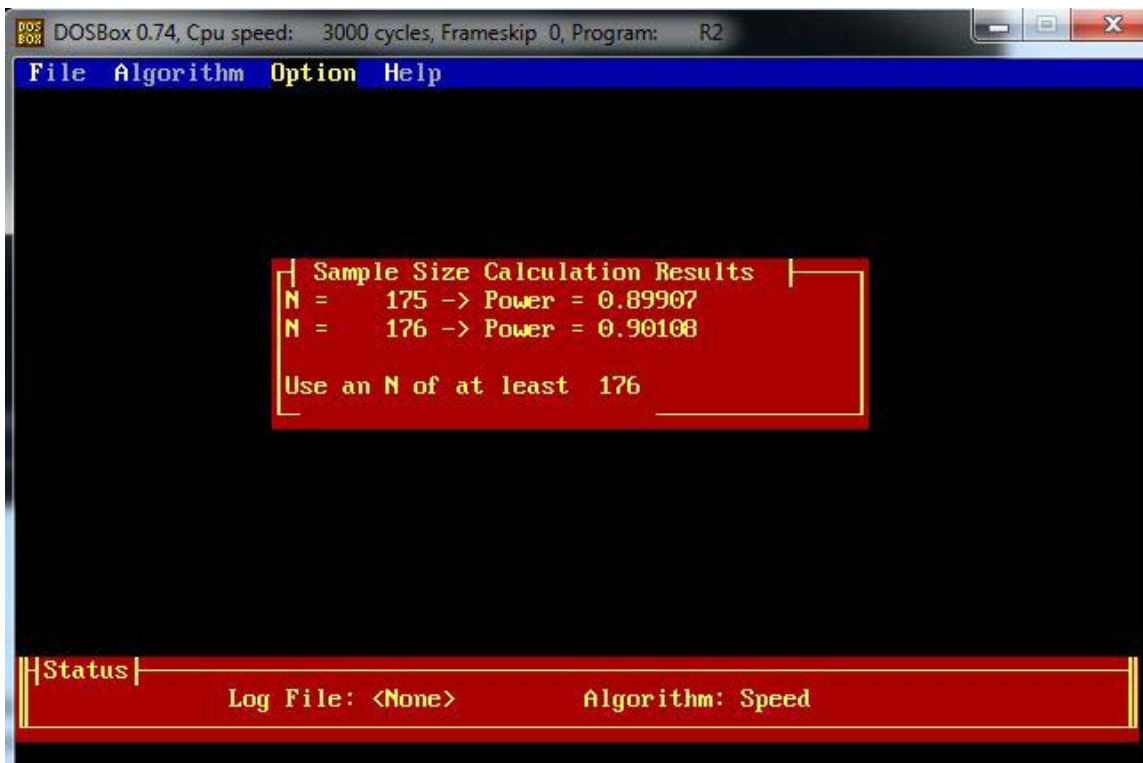
2. **Power Calculation for Tests of Significance of the Squared Multiple Correlation**

R2 contains some powerful utilities for power calculations in multiple regression. Power of the test $P^2 = 0$ can be computed for any alternative. Suppose, for example, you were anticipating a low value of P^2 (say .10) and were planning a study with 7 predictors and 50 subjects. A natural question to ask is whether power would be adequate in this situation. After activating the "Power Analysis" submenu from the Options main menu,

enter the following values: $N = 50$, $K = 8$, $\alpha = 0.05$, $P^2 = 0.10$. You should obtain a calculated power of .29283.

3. Calculating the Sample Size Necessary to Obtain a Desired Power Level

The above calculation shows that power is simply inadequate. The question then becomes something like, "how large should N be in order to have a power of, say, .9." R2 allows direct calculation of the answer to this question. (The program does require the requested power to be greater than .50.). The program iterates until it find the values of N which bracket the requested power. To estimate the required sample size in the current situation, activate the "Sample Size Calculation" submenu from the "Options" main menu, and enter the values $P^2 = 0.10$, $K = 8$, $\alpha = 0.05$, Desired power = .90. After you start the calculations, the program will iterate until it finds two values (if possible) greater than $K + 1$ which bracket the desired level of power. It will report back to you as follows:

4. Testing Hypotheses of the Form $P^2 = a$, for Values of a Other Than Zero.

This is done rather routinely by calculating a confidence interval, and seeing whether the interval includes or excludes *a*. If a confidence interval at the $1-\alpha$ confidence level *does not* include *a*, then the two tailed hypothesis test at the *a* level will reject the null hypotheis that $P^2 = a$. One-tailed tests can be performed in an analogous way using the confidence bounds.

For example, suppose you observe an R^2 value of .5, and you wished to test the hypothesis $P^2 \leq 0.2$ against the alternative $P^2 > 0.2$. Suppose $N = 134, K = 8$. The calculations shown below indicate that a 95% lower bound for for P^2 exceeds .2, indicating that the null hypothesis may be rejected at the .05 level.

Here are the input values:

```
 File  Algorithm  Option  Help

                  ┌Confidence Interval Input Values┐
                   Number of observations:                   134
                   K (Number of variables, including criterion):  8
                   R Squared:                                 .5
                   Confidence Level:                          .95
                   Go <Begin Computing>



╫Status┠
              Log File: <None>          Algorithm: Speed
```

Here are the results:

5.   Probability Integral Calculations

R2 can calculate the exact probability integral for the distribution of R^2. This is the probability that a randomly sampled squared multiple R will be less than or equal to a given value. Simply activate the menu, fill in the required quantities, and activate the "<Go>" option.

Aspects of Program Operation

1. Error Messages.

R2 tries to identify errors in input values as quickly as possible. When errors occur, the program will flash an error message, then allow you to re-enter the incorrect value. Most error messages will clearly define the problem area for you. Practice using the error message facility by trying some illegal values. Try entering a negative *N* or a squared multiple correlation which is greater than 1 to observe how the system reacts.

2.   Beeps

If, while operating the menu system, you press an incorrect key, the program will beep. This can be somewhat annoying, but after you have gained minimal experience with the program, you will make very few errors. Most "beeps" tend to occur when the user attempts to enter a value for a variable (often N) before either pressing the key letter for the variable or hitting <Enter> to activate the data entry mode.

3.   Extended Calculation Times

Some operations, expecially confidence interval calculations, can take very little time for small to moderate N and K values, but can take several minutes if the squared multiple correlations are above .9. As an indication of how long you can expect a calculation to take, R2 prints iteration numbers during calculations involving the fast algorithm. During the slower more accurate algorithm, the program adds, *for each iteration*, a row of dots. For each iteration there will be approximately 20 dots. For any given calculation, the number of dots will remain very constant. (*Note: these indicators may not appear on some machines using nonstandard graphics adaptors.)*

R2 prefaces iterative calculations involving the slower algorithm with a preliminary calculation using the faster algorithm. Because the faster algorithm almost always gives results which are very accurate, the calculations involving the slower algorithm usually require no more than 3 iterations, often only one or two.

Some calculations can take several minutes, especially on older computers with 8088 or 80286 CPUs without a math coprocessor. Fortunately, the examples for which R2 takes the longest computation time (large numbers of variables, high multiple correlation values) tend to occur much less frequently than calculations which can be performed quickly.

Problem Reports

We have attempted to test R2 on a broad range of problems, and we will continue our testing program in the future. If you encounter problems with R2, or have suggestions for improving the program, please communicate them to the senior author.

Surface mail: James H. Steiger
Department of Psychology & Human Development
Peabody College
Vanderbilt University
Nashville, TN 37203

Email: james.h.steiger@vanderbilt.edu

References

Cox, D.R., & Hinckley, D.V. (1974). *Theoretical Statistics.* New York: Chapman & Hall.

Lee, Y-S. (1971). Some results on the sampling distribution of the multiple correlation coefficient. *Journal of the Royal Statistical Society, Series B, 33,* 117-130.

Lee, Y-S. (1972). Tables of the upper percentage points of the multiple correlation coefficient. *Biometrika, 59,* 175-189.

Steiger, J.H., & Fouladi, R.T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.) *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates. (May be download from the following the "Publications" section of http://www.statpower.net in the "Publications" section under Book Chapters.