FISH

Friendly Introductory Statistics Help (Version 3.0.3)

Presented at the Joint Statistical Meetings, August 2004, Toronto, Canada

Statisfaction Guaranteed (with 95% Confidence)

Contact: brooksg@ohio.edu

Web Site URL: http://oak.cats.ohiou.edu/~brooksg/fish.htm

Instructor's Manual

Prepared by Holly Raffle

Table of Contents

1 Introduction and Features

Friendly Introductory Statistics Help (FISH) is designed as a free, easy-to-use program for use by both students and instructors. The software's features make it an excellent teaching tool for secondary school, undergraduate and graduate level introductory statistics courses. For teaching purposes, FISH generates sample sets of data from a population defined by the user with a random data generator. This allows instructors to provide tangible demonstrations of abstract concepts such as the central limit theorem and sampling error. Because of FISH also allows users to manually enter data sets, it is appropriate for students to use as a self-paced tutorial package. The step-by-step presentation and the easily accessible "help" buttons allow students to use FISH as an independent study of basic statistical concepts.

FISH allows users to:

<u>Analyze One Variable.</u> This option calculates measures of central tendency, deviation scores, squared deviation scores, sum of the squared deviation scores, variance, range, standard deviation, z-scores, standard error of the mean, and the confidence interval of the mean.

Within this option, the user is able to toggle between the above calculations computed with (N – 1) for sample statistics and (N) for population parameters. Users are also able to toggle between several options for calculating confidence intervals. Using the z-distribution, the following confidence intervals can be calculated: 68% (z = 1.00), 90% (z = 1.645), 95% (z = 1.96), and 99% (z = 2.33). Using the t-distribution with n – 1 degrees of freedom, confidence intervals can be calculated at the following alpha levels: $\alpha = .001$, $\alpha = .01$, $\alpha = .05$, and $\alpha = .10$.

This option also allows the user to view a graphical representation of the results. Users are able to access frequency tables (with toggles allowing raw scores, z-scores, and both raw and z-scores to be shown) and grouped frequency tables.

When the user chooses to generate random data, the histogram feature is enhanced to allow the user to view the sampling distribution of the mean. From this screen, the user can generate additional samples in multiples of 1, 10, 100 or 10,000. These sample means are then added to the existing sampling distribution of the mean.

Analyze Two Variables. This option calculates the means, standard deviations, deviation scores, and cross products for both variables. Additionally, the unstandardized correlation coefficient (covariance), correlation coefficient (Pearson's r), coefficient of determination (R^2), regression slope and Y-intercept are calculated.

As in the Analyze One Variable option, the user is able to toggle between the above calculations computed with (N – 1) for sample statistics and (N) for population parameters.

This option also allows the user to view a graphical representation of the results. Users are able to access scatterplots (with toggles allowing raw scores, z-scores, and both raw and z-scores to be shown) and have the option to show or hide the mean lines and the regression line. The histogram option also shows the regression equation and the values for the correlation coefficient and the coefficient of determination. Additionally, the user has access to observed, predicted, and residual values, squared deviation scores, model sum of squares ($(\hat{Y} - \bar{Y})^2$), as well as calculations of the standard error of the estimate, standard error of the regression slope, and the 95% confidence interval for the regression slope.

When the user chooses to generate random data, the scatterplot feature is enhanced to allow users to view the sampling distribution of the regression coefficient. From this screen, the user can generate additional samples in multiples of 1, 10, 100 or 10,000. These sample regression coefficients are then added to the existing sampling distribution of the regression slope.

2 Installation and Setup

FISH operates in any Windows environment. To install FISH:

1. Insert the CD-ROM into the proper drive.
2. View the contents of the CD-ROM using Windows Explore.
3. Click on the FISH icon. The program will launch and you are ready to begin.

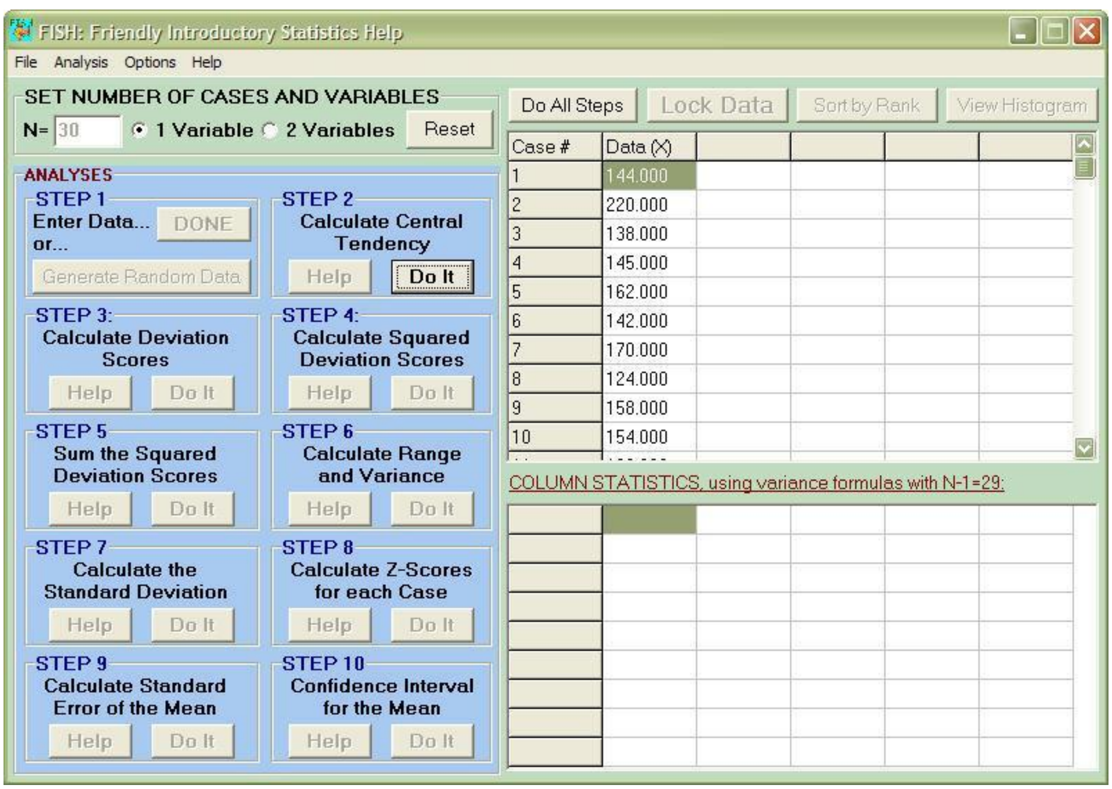
3 Using the Instructor's Manual

This Instructor's Manual is designed to complement the FISH User's Guide. The manual is intended to present learning activities and pedagogical pearls; it is not intended to be a comprehensive guide for using FISH. For software support and step-by-step instructions, please refer to the FISH User's Guide.

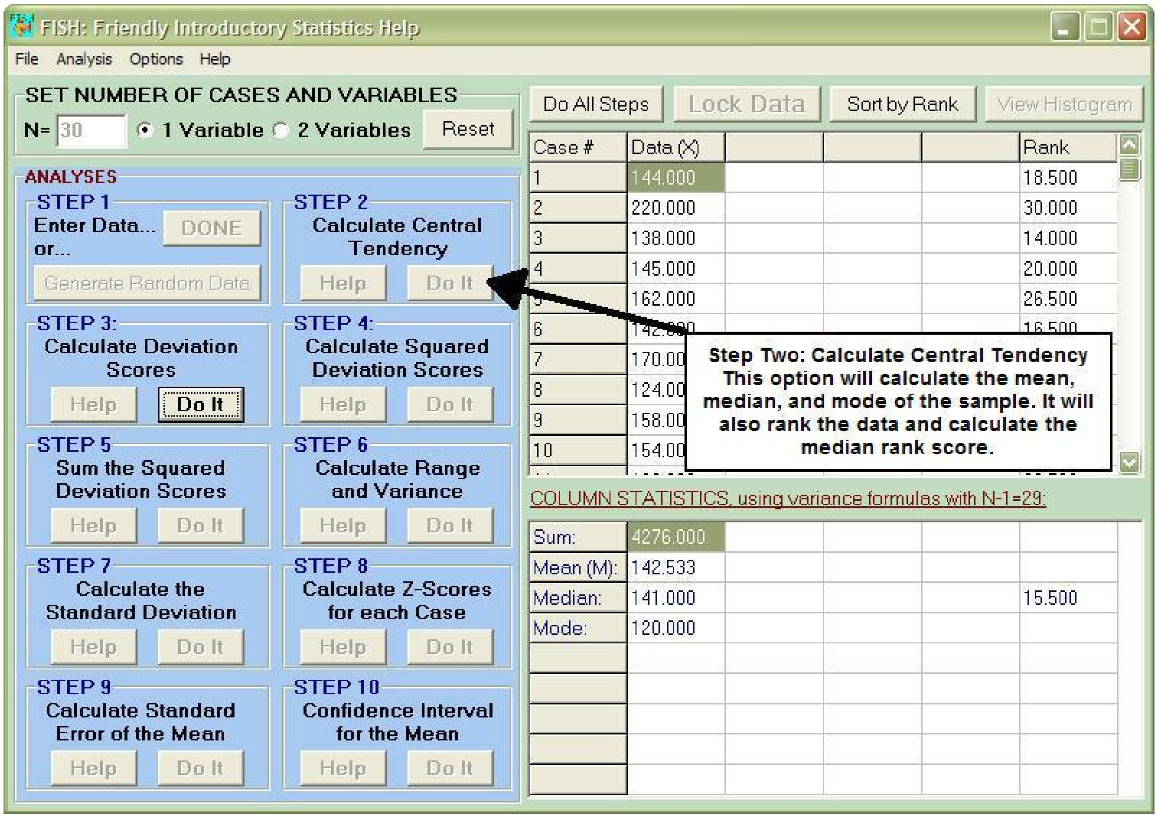
4 Analyzing One Variable Using a FISH Data File

To illustrate uses for FISH in the classroom setting, a data file has been included with this Instructor's Manual. The data (SBP FISH_data) can be found on the CD-ROM. The data set was taken from Lemeshow, 2003 (*Applied Regression Analysis Course Notes*, 2003 Summer Program in Applied Statistical Methods, p. 4). The data includes observations on systolic blood pressure for a sample of 30 individuals.

Open the file "SBP FISH_data" and import the data into FISH. The data will appear on the screen as follows:



Step Two: Calculate Central Tendency



Mean:

The mean is the most common measure of central tendency. It is also called the arithmetic average or the arithmetic mean. The population mean, μ, is calculated:

$$\mu = \frac{\sum X}{N}$$

where N indicates the population size.

The sample mean, also called the estimate mean, $\overline{X}$, is calculated:

$$\overline{X} = \frac{\sum X}{n}$$

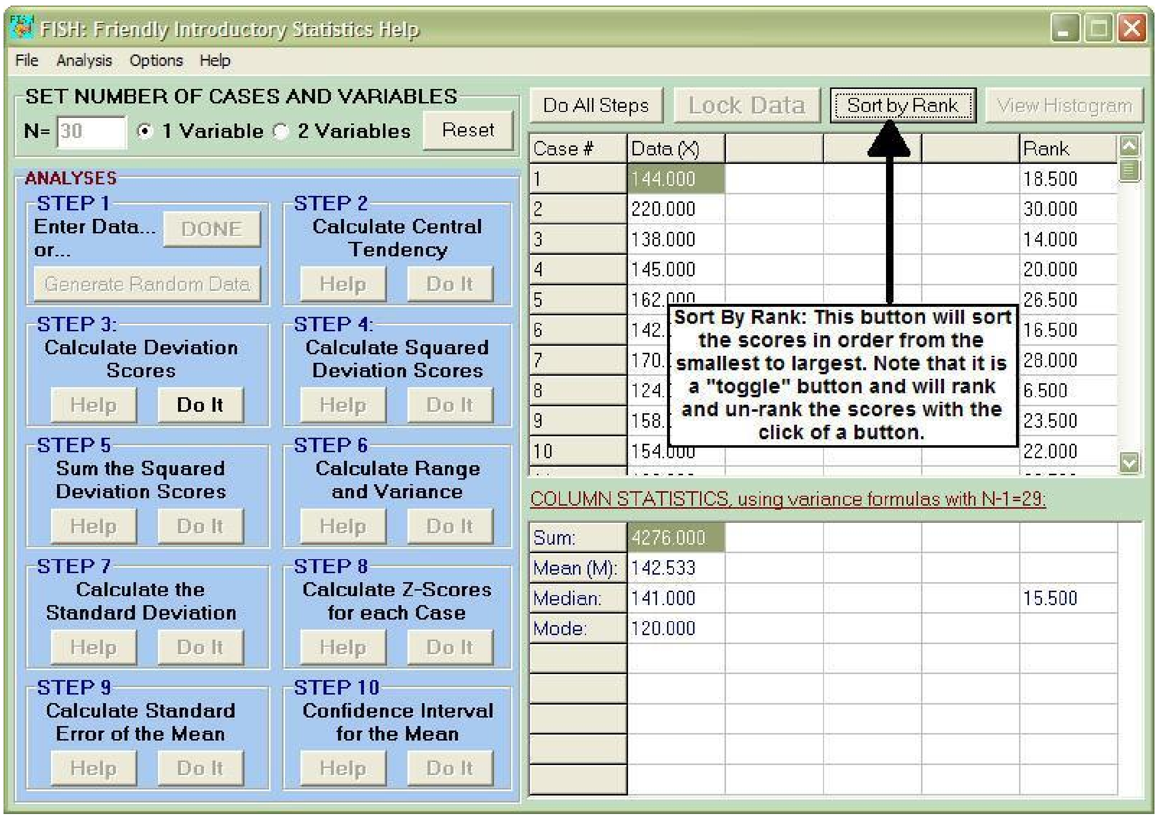
where n indicates the sample size.

Using the systolic blood pressure (SBP) data, students can confirm for themselves the calculation by using the sum score provided by FISH and dividing it by the sample size.

Median:

The median is the score that lies in the middle of a distribution when the scores are arranged in increasing order. The median is also called the 50th percentile or the score at or below which 50% of the scores in a distribution fall.

To illustrate this concept using FISH, utilize the “Sort By Rank” Option.

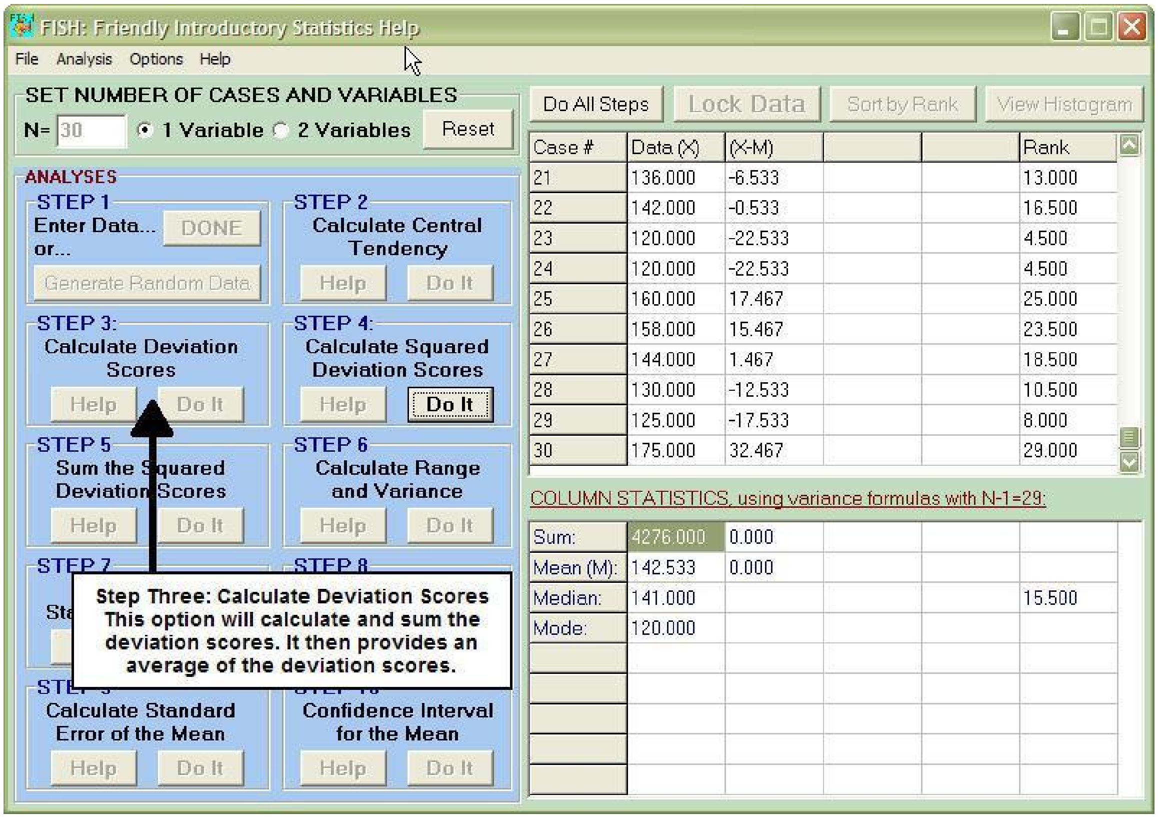


Mode:

The mode of a distribution of scores is the score that occurs most frequently. This is also easily illustrated by using the “Sort By Rank” Option. When the scores are ranked, it is easy to find the most frequent score.

Note: The current version of FISH (2.5.8) does not have the ability to determine if a distribution is bimodal or multimodal. It is important to inform students of this limitation, and to make corrections accordingly.

Step Three: Calculate Deviation Scores



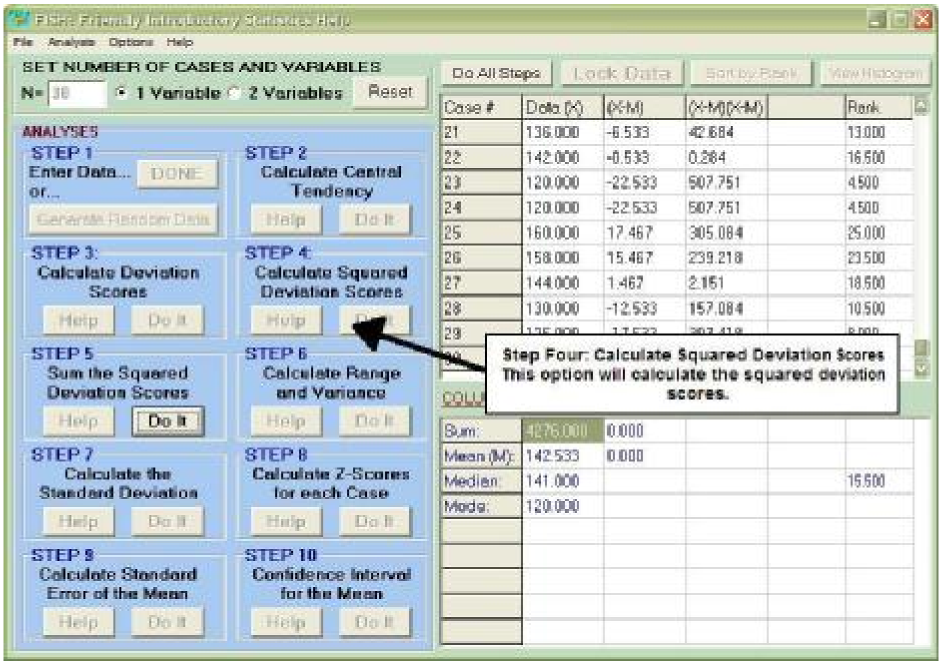
FISH makes it easy for students to visualize the following property of the mean: The sum of the deviation scores about the mean equals 0.

$$\sum\left(X-\bar{X}\right)=0$$

Because the sum of the deviation scores equals 0, the mean of the deviation scores is clearly 0.

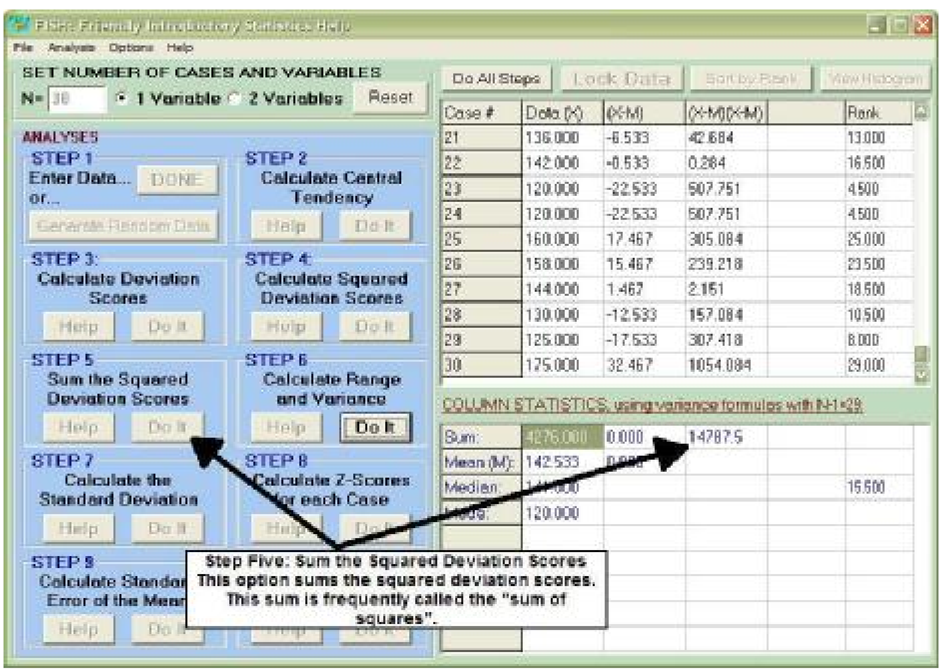
Students can confirm this property for themselves by verifying the calculations computed by FISH.

Step Four: Calculate Squared Deviation Scores



Because the sum of the deviation scores is 0, it is not very interesting or useful. Therefore, the deviation scores are squared to provide better insight into the data.

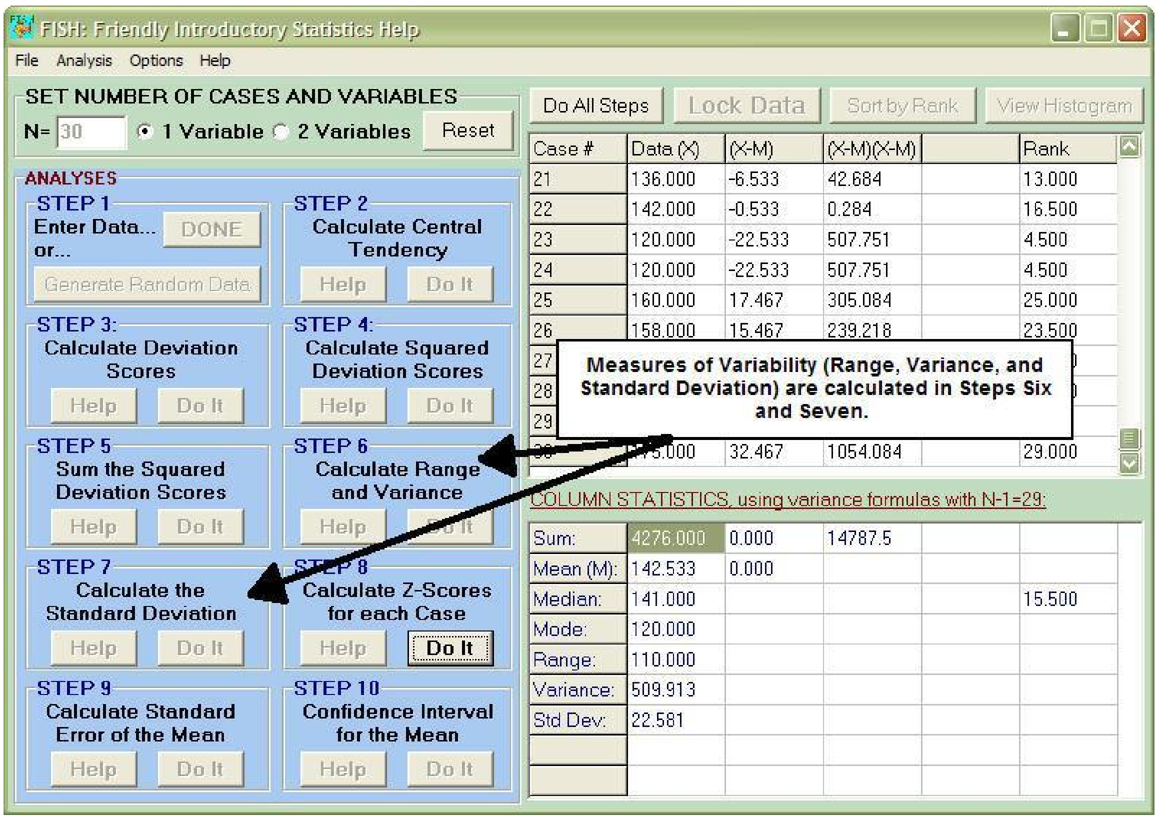
Step Five: Sum the Squared Deviation Scores



The transition from the sum of the deviation scores to the sum of the squared deviation scores is a good way to introduce the concept of the *least-squares estimate* approach that is used in more advanced statistical procedures such as ANOVA.

The sum of the squared deviation scores or “sum of squares” is smaller than the sum of the squared deviation scores about any other number. The least-squares estimate is named such because the mean is used to minimize the squares of the deviation scores compared to the squares of the deviation of the scores from any other value.

Steps Six and Seven: Calculate Range, Variance, and Standard Deviation



Measures of variability are useful to describe the dispersion of scores within a distribution.

<u>Range:</u>

The range is simply the measure of the distance between the largest and smallest score within a distribution. FISH computes the range using this formula:

$$\text{Range} = (\text{largest score} - \text{smallest score})$$

The *inclusive range* is sometimes reported (and called *range* at times). The inclusive range is calculated using this formula:

Inclusive range = (largest score – smallest score) + 1

The "1" is added to indicate that the distance being measured is the upper limit of the largest score and the lower limit of the smallest score. If the scores are whole numbers, the largest score will be .5 below the upper limit and the smallest score will be .5 above the lower limit (.5 + .5 = 1).

Although the range and inclusive range are easy to calculate, they can be less than useful for two reasons: 1) The range is dependent on the two most extreme scores and 2) The range will likely increase as sample size increases.

<u>Variance and Standard Deviation:</u>
As discussed previously, deviation scores are the distance from each raw score to the mean. Recall that a property of the mean is that the sum of the deviation scores from the mean is 0. This property made the sum and the average of the deviation scores neither interesting nor useful. To compensate for this property of the mean, there are two obvious choices: 1) Take the absolute value of the deviation scores or 2) Square the deviation scores.

The *mean absolute deviation* is the sum of the absolute values of the deviations from the mean. This is seldom used in statistics. Instead, the terms are squared. The *sum of squares* is the sum of the squared deviation scores. The sum of squares is the preferred method and is used to calculate variance and standard deviation. The sum of squares is also used in many other statistical procedures.

The *variance* of a distribution is not a widely reported measure of variability. However, it is very useful in other analyses. The variance is the average of the sum of squares. When a researcher has measured all possible scores in the population, the population variance is calculated in the following manner:

$$\sigma^2 = \frac{\sum (X - m)^2}{N}$$

To provide an unbiased estimate of the population parameter σ^2, the following formula is used to calculate the sample variance:

$$s^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$

Note that FISH allows users to “toggle” from calculations using N and N-1 at the touch of a button.

The *standard deviation* of a distribution is the square root of the average sum of squares. (Recall that the average sum of squares is the variance.) Squaring the deviation scores and then taking the square root preserves the units (in the sample SBP data, mmHg) of the original scores. Like the variance, there is a formula for the population standard deviation and the sample standard deviation. The population standard deviation is calculated with the following formula:

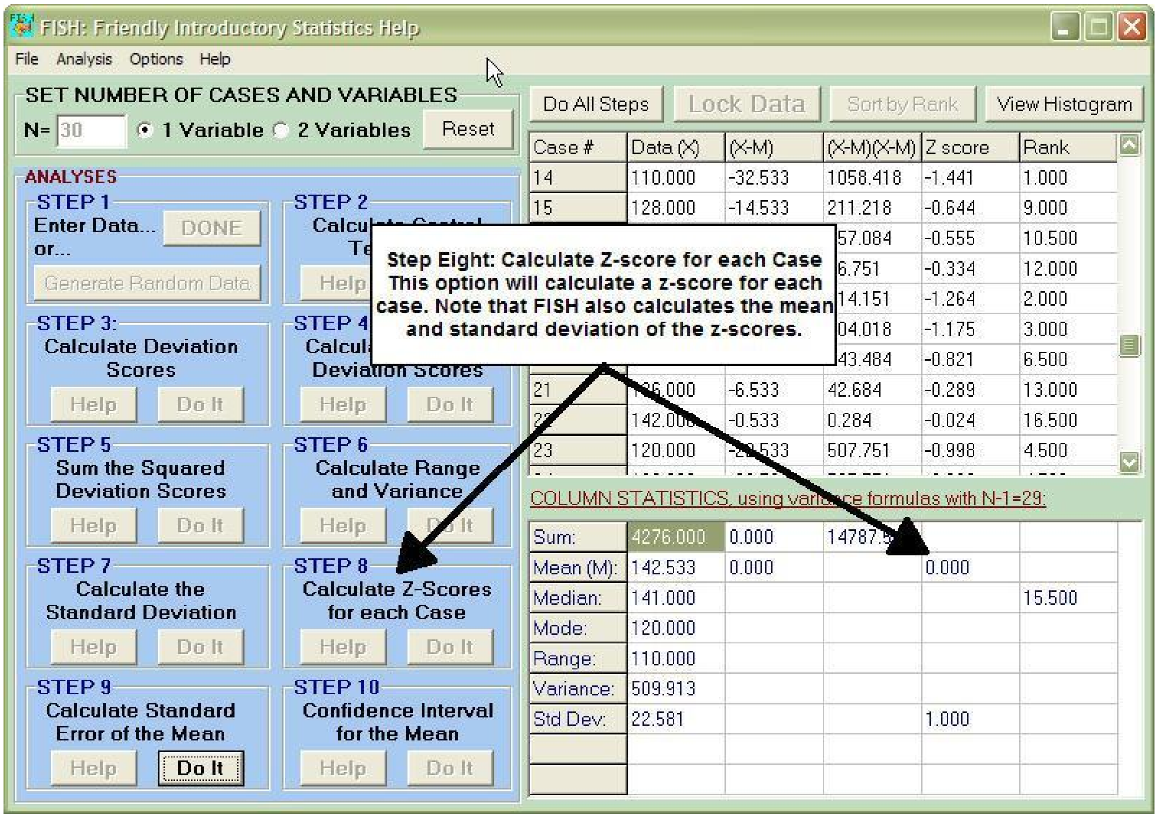
$$\sigma = \sqrt{\frac{\sum (X - m)^2}{N}}$$

The sample standard deviation is calculated:

$$s = \sqrt{\frac{\sum (X - \overline{X})^2}{N - 1}}$$

Recall that FISH allows users to “toggle” from calculations using N and N-1 at the touch of a button

Step Eight: Calculate Z-Score for Each Case



A *Standard Score* (z-score) relates a raw score to the mean and standard deviation its' distribution. A z-score represents the number of standard deviations a score is above (for a positive z) or below (for a negative z) the mean of a distribution. For a population, z-scores are calculated using this formula:

$$z = \frac{X - \overline{X}}{s}$$

For a sample, the following formula is used:

$$z = \frac{X - m}{s}$$

The *z-distribution* has a mean = 0 and a standard deviation = 1. Using the distribution of scores provided with the SBP data, this can be confirmed using FISH calculations by looking at the "column statistics" for the z-scores.

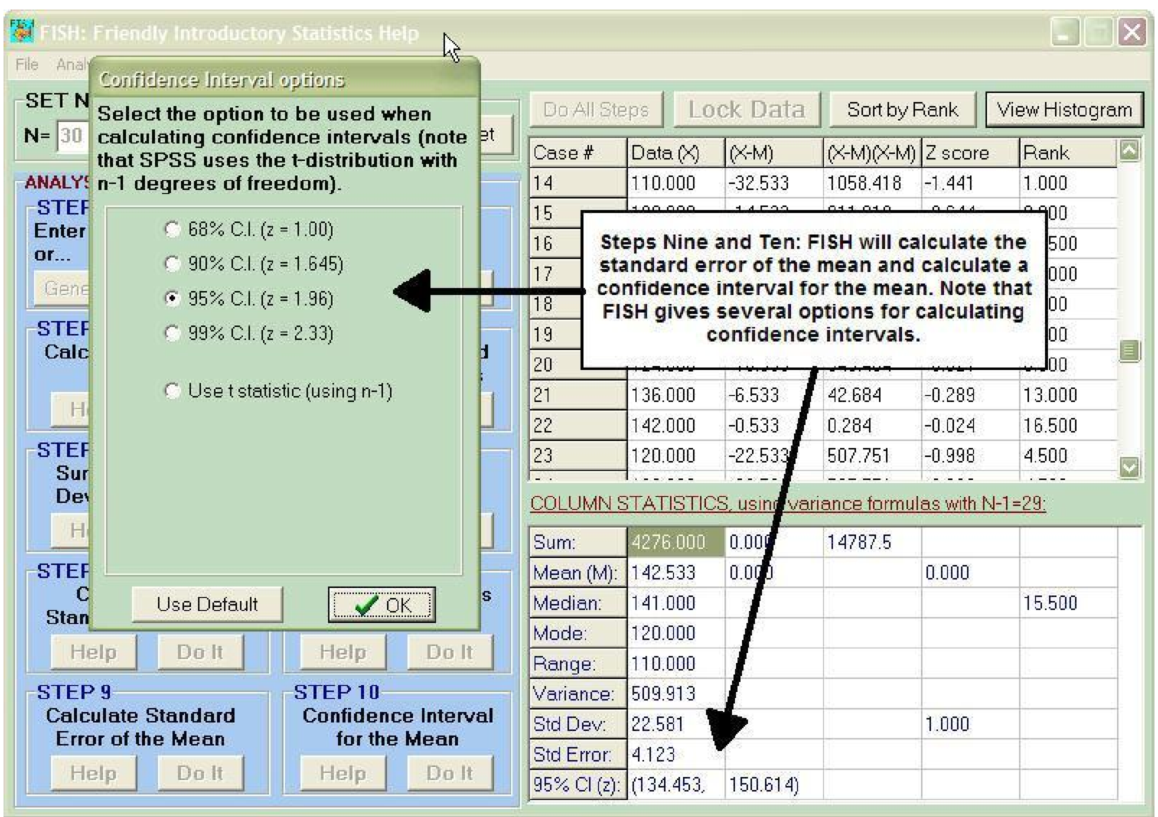
Why use z-scores?

1. Because the mean of a z-distribution is 0, the *sign* of the z-score is informative. By inspection, you can quickly determine if a score is above of below the mean. If the z-score is positive, it lies above the mean; if the z-score is negative, it lies below the mean.

2. Because the standard deviation of a z-distribution is 1, the *magnitude* of the z-score is also informative. When the distribution of scores is approximately normal, 2/3 of the scores fall between +1 and -1, and 95% of the scores fall between -2 and +2.

3. Z-scores permit comparisons across distributions with different means and standard deviations. This is only true if the distributions being compared are reasonably equivalent.

Steps Nine and Ten: Calculate Standard Error of the Mean and Confidence Interval for the Mean



Up to this point, we have considered a distribution of raw scores, the SBP data. The standard error of the mean and the confidence interval of the mean are based upon a *theoretical* distribution: the sampling distribution. Note that a sampling distribution is different from a distribution of raw scores. A sampling distribution is based on the idea of repeated sampling: drawing an infinite number of samples from a population. The mean of all scores in a sampling distribution is called the *expected value* of the statistic (in our case the statistic is the mean) for that particular distribution. The standard deviation of a sampling distribution of a statistic is called the *standard error* of the statistic.

Because the statistic we are interested in is the mean, FISH is calculating the *standard error of the mean*. When the population standard deviation (σ) is known, the standard error of the mean is calculated in the following manner:

$$s_{\bar{X}} = \frac{s}{\sqrt{N}}$$

When the population standard deviation is unknown, the standard error of the mean is calculated using this formula:

$$s_{\bar{X}} = \frac{s}{\sqrt{N}}$$

A *confidence interval* is a range of scores that provides a specific probability that the interval will contain the population parameter. In our case, we are estimating the *confidence interval of the mean*; therefore, the mean is the population parameter. If you had the ability to draw 100 samples from the population and calculate the 95% confidence interval for each sample (100 total intervals), approximately 95 of the confidence intervals would contain the population parameter (and 5 will not).

Confidence intervals can be calculated using both the z-distribution and the t-distribution. When σ is known and the sample size is large (> 120), confidence intervals are calculated using the z-distribution:

68% Confidence Interval:	$\bar{X} \pm 1.00 s_{\bar{X}}$
90% Confidence Interval:	$\bar{X} \pm 1.645 s_{\bar{X}}$
95% Confidence Interval:	$\bar{X} \pm 1.96 s_{\bar{X}}$
99% Confidence Interval:	$\bar{X} \pm 2.33 s_{\bar{X}}$

When σ is unknown and the sample size is small (< 120), confidence intervals are calculated using the t-distribution:

$$X \pm [t_a(N-1)]S_{\bar{X}}$$

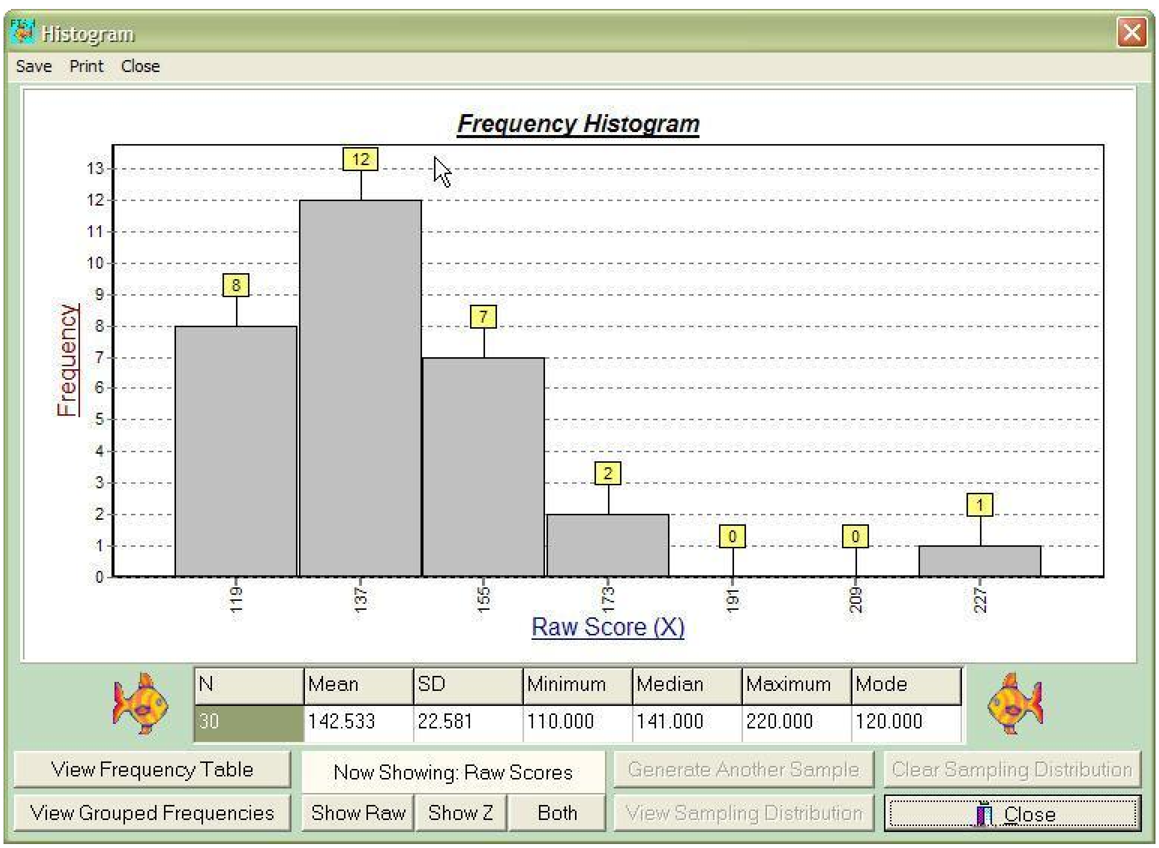
where $t_a(N-1)$ is the tabulated *t* value with N-1 degrees of freedom at the desired α.

It is important to note that standard errors and confidence intervals can be computed for other statistics such as the regression coefficient (slope) and the correlation coefficient.

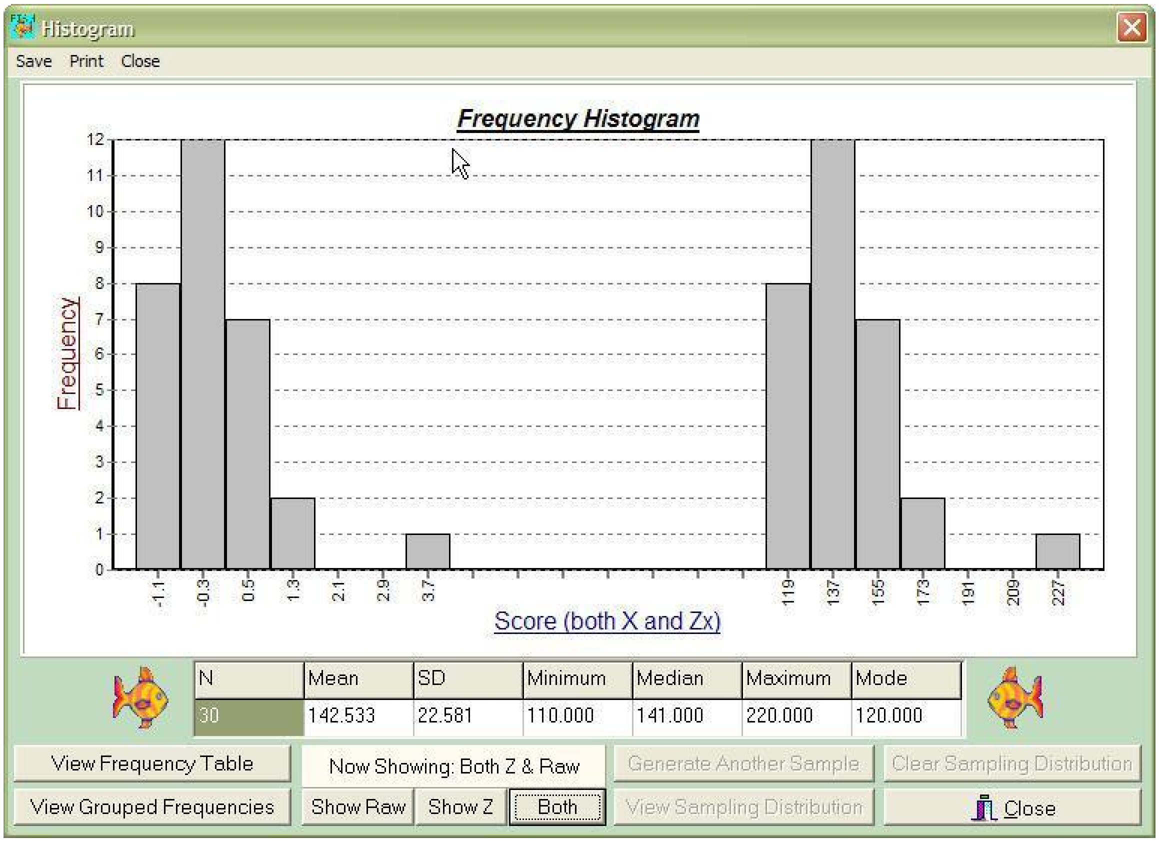
Recall that FISH allows users to "toggle" between several confidence interval calculations with the click of a button.

Histogram Feature

It is important to examine a distribution of data graphically. Using the histogram option, FISH allows users view the distribution of the data.



Using the "Now Showing" buttons, FISH allows users to "toggle" between the following options: Show Raw Scores, Show Z-Scores, and Show Both. The *Show Both* option offers a concrete way for students to verify that the distribution of raw scores and the distribution of z-scores have the same shape.

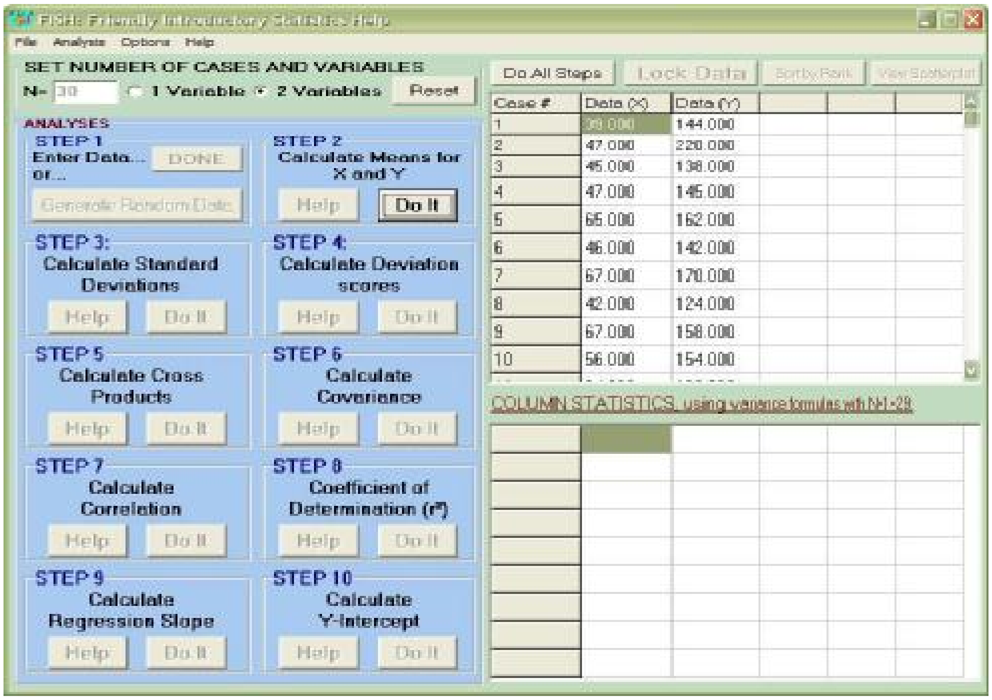


Please refer to the User's Guide for a more thorough examination of FISH's graphics capabilities.

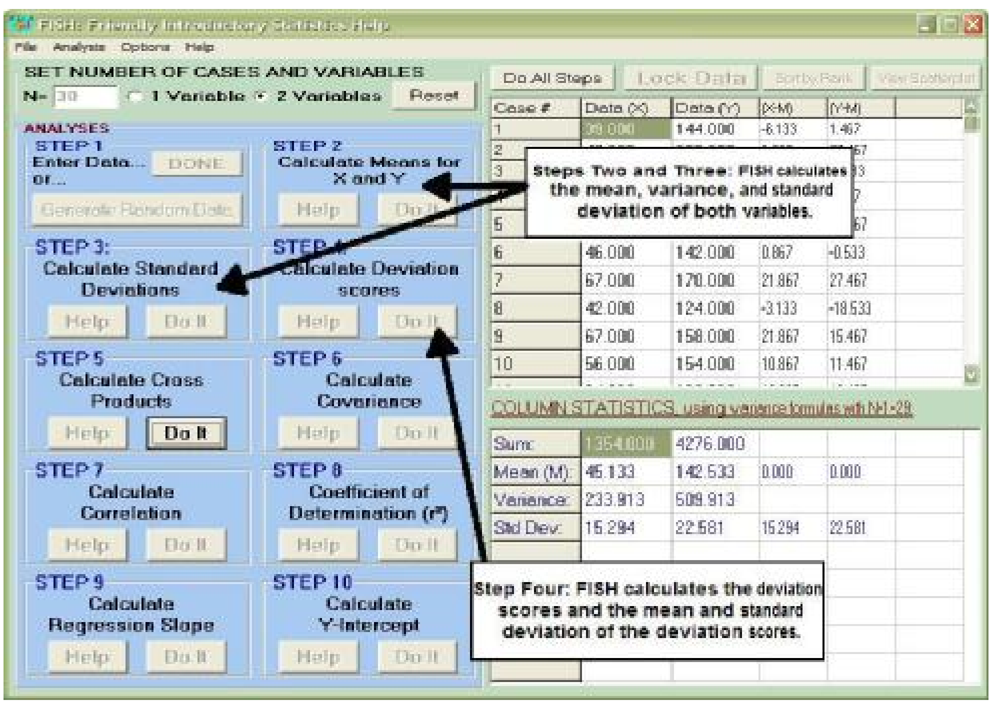
5 Analyzing Two Variables Using a FISH Data File

To illustrate uses for FISH in the classroom setting, a data file has been included with this Instructor's Manual. The data (SBP & AGE FISH_data) can be found on the CD-ROM. The data set was taken from Lemeshow, 2003 (*Applied Regression Analysis Course Notes*, 2003 Summer Program in Applied Statistical Methods, p. 4). The data includes observations on age and systolic blood pressure for a sample of 30 individuals.

Open the file “SBP & AGE FISH_data” and import the data into FISH. The data will appear on the screen as follows:

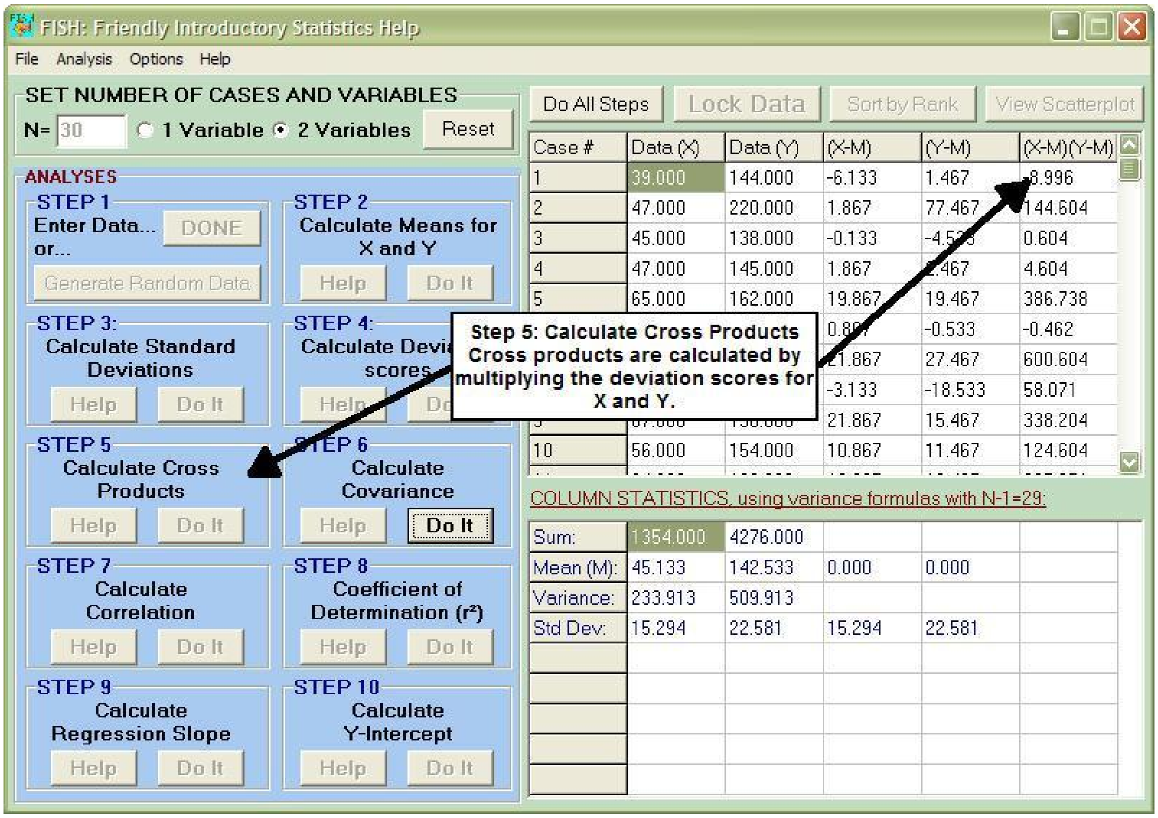


Steps Two, Three, and Four: Calculate Means for X and Y, Calculate Standard Deviations, and Calculate Deviation Scores



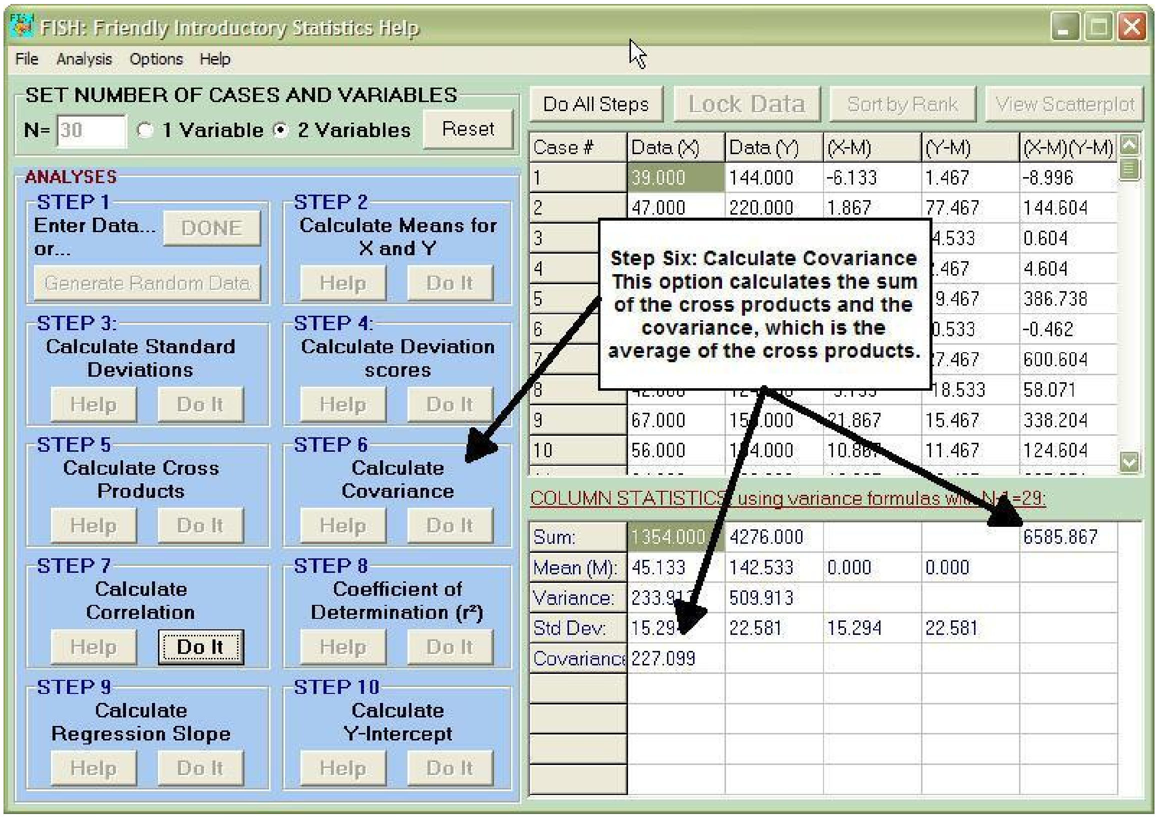
Please see pages 5-9 for information on calculating and interpreting measures of central tendency and variability.

Step Five: Calculate Cross Products



Just as the deviation scores are important in the sum of squares, the cross products are important in calculating the sum of the cross products.

Step Six: Calculate Covariance



The *covariance* is the average sum of the cross products. It is also referred to as the average sum of the combined differences. Covariance is calculated with this formula:

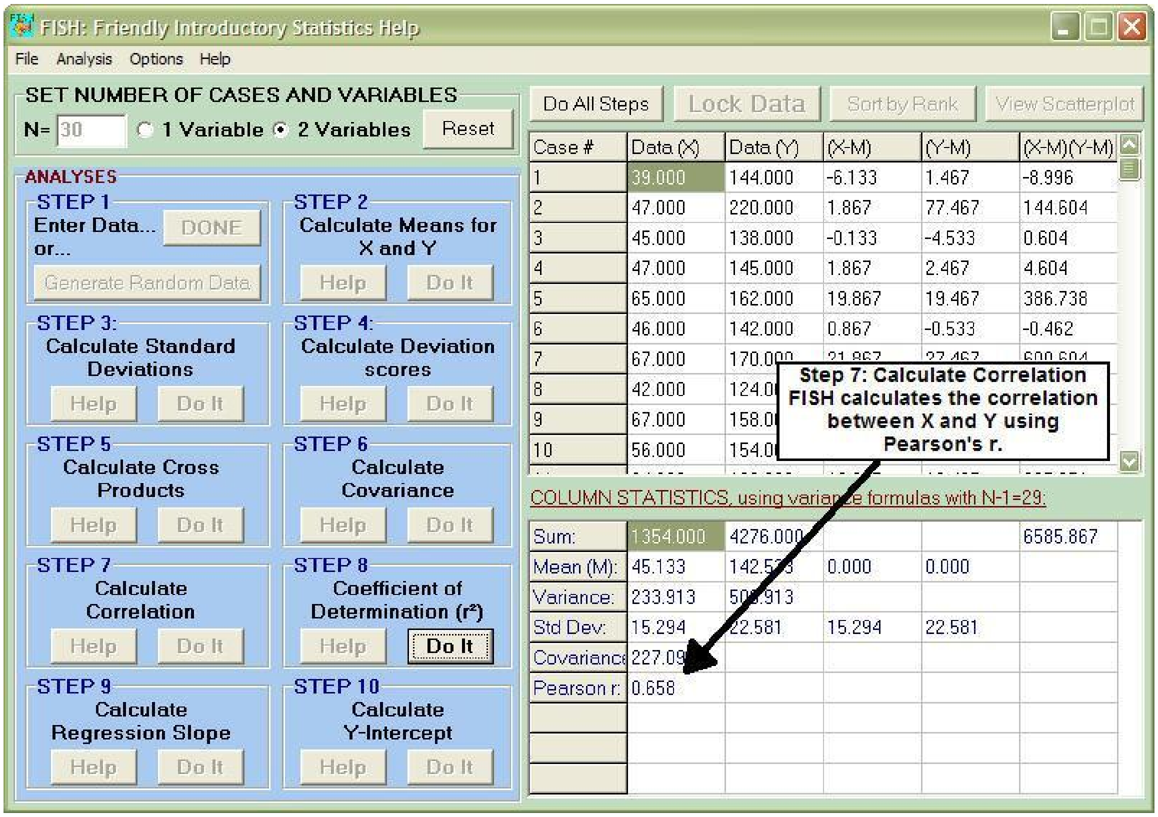
$$\text{cov}_{(x,y)} = \frac{\sum (X - \bar{X}) - (Y - \bar{Y})}{N - 1}$$

Covariance is also known as the *unstandardized correlation coefficient*. Covariance can be used to assess whether two variables are related. A *positive covariance* indicates that as one variable deviates from the mean, the other variable deviates in the same direction (e.g.: as X increases, Y also increases). A *negative covariance* indicates that as one variable deviates from the mean, the other variable deviates from the mean in the opposite direction (e.g.: as X increases, Y decreases).

Because covariance is not standardized, it is scale dependent. That is, you can not compare the covariance of two distributions in an objective fashion unless both distributions are measured in the same units.

Step Seven: Calculate Correlation

To standardize the covariance term, the standard deviations of X and Y are used. The *standardized* correlation coefficient is known as the Pearson Product-Moment Correlation Coefficient or Pearson's *r*.



Pearson's *r* is calculated conceptually with the following formula:

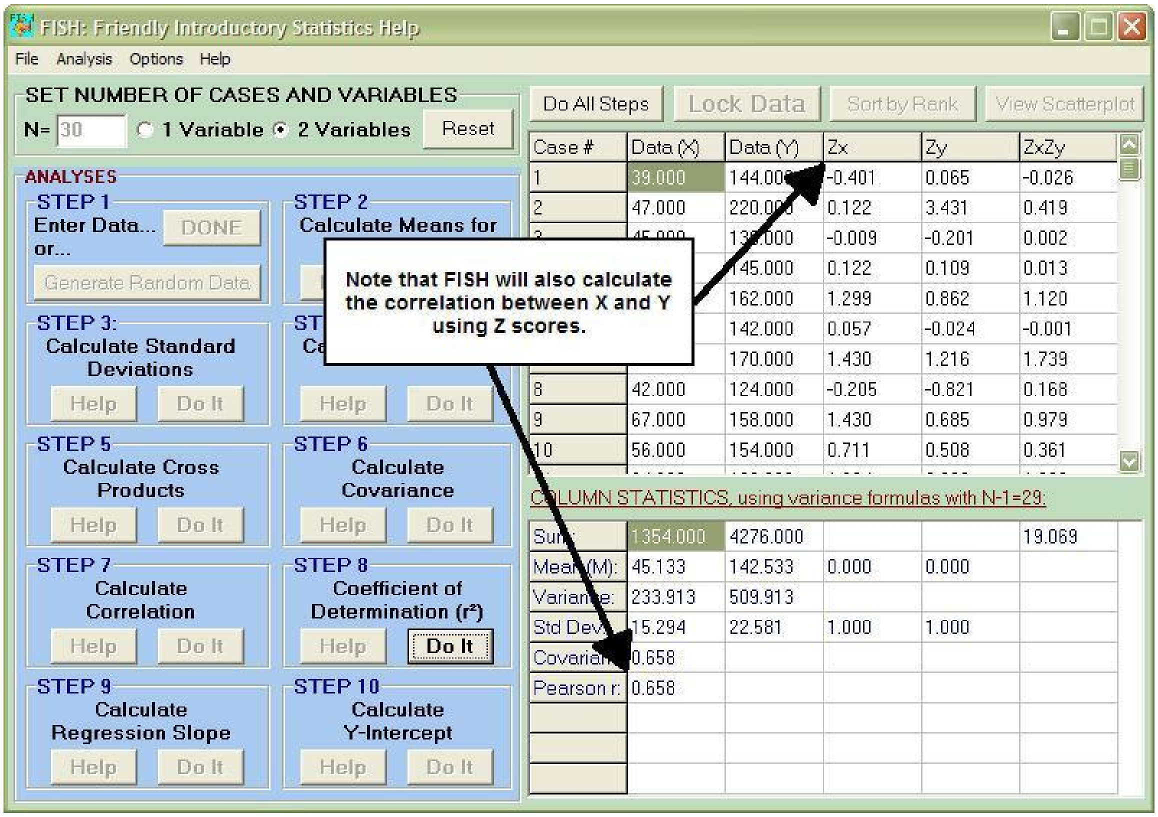
$$r = \frac{\mathrm{cov}_{xy}}{s_x * s_y}$$

Computationally, the formula for Pearson's *r* is:

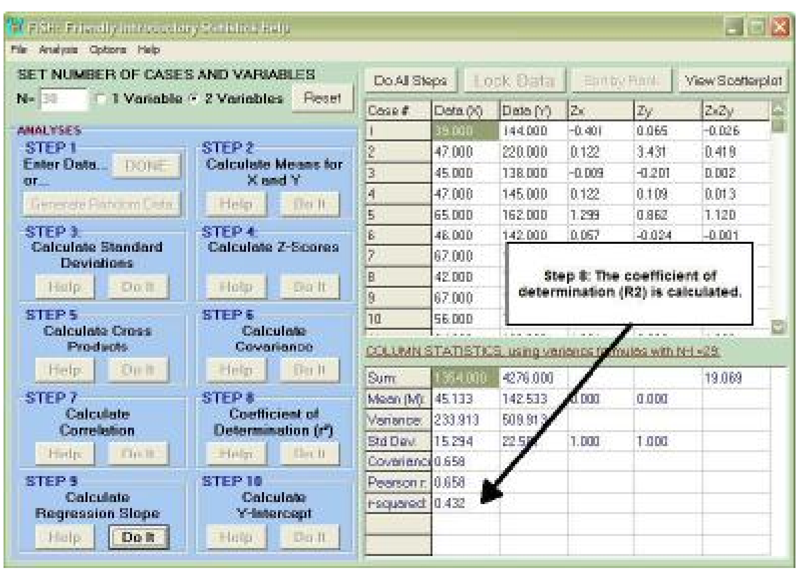
$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{(N-1)s_x * s_y}$$

By standardizing the covariance, values for *r* range from +1 to -1. When *r* = +1, X and Y are perfectly, positively correlated. As one variable increases, the other variable increases proportionally. When *r* = -1, X and Y are perfectly, negatively correlated. When *r* = 0, there is no linear relationship between X and Y.

Note that FISH will calculate *r* using either raw scores or z-scores. Users may toggle between the two options. This will allow users verify two items: 1) When using standard scores, the covariance term = correlation coefficient and 2) The correlation coefficient is the same when calculated with z-scores.



Step Eight: Coefficient of Determination (r^2)



The *coefficient of determination* (r^2) is a measure of the strength of the relationship between two variables. The coefficient of determination is sometimes referred to as the proportion of variance on one variable predictable from or attributable to the other variable. The coefficient of determination is calculated by simply squaring the correlation coefficient:

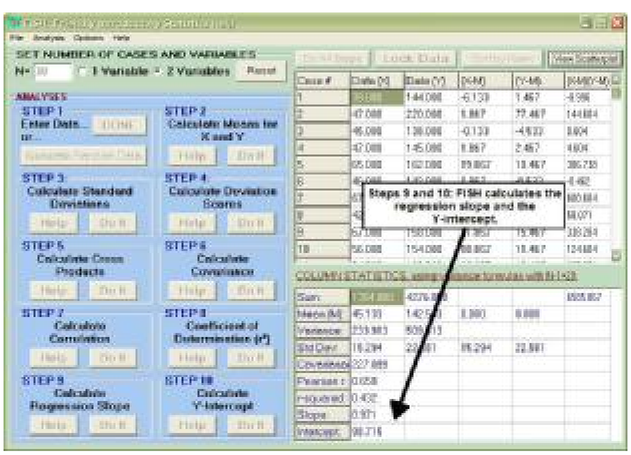
$$r = \left(\frac{\sum \left(X - \overline{X}\right)\left(Y - \overline{Y}\right)}{(N-1)s_x * s_y} \right)^2$$

If r^2 is multiplied by 100, it can be thought of as a measure of the percentage of variance explained by correlation between the two variables. In the example FISH data set, $r^2 = .432$. Therefore, the correlation with age (X) accounts for 43.2% of the variance in systolic blood pressure (Y).

Steps Nine and Ten: Calculate Regression Slope and Calculate Y-Intercept

Regression is a statistical procedure which fits a predictive model to the data and uses that model to predict values of the dependent variable from one or more independent variables. FISH provides an example of *simple regression*, the prediction of an outcome (dependent variable) from a single predictor (independent variable).

Regression models are a linear representation of the distribution of scores. Regression models attempt to summarize a distribution of scores with a straight line. There are many lines which can summarize the general trend of a distribution. Regression employs the *method of least squares* (see page 9), a mathematical technique to establish the line that BEST fits the distribution of scores.



Review of Straight Lines:
Any line can be drawn if two things are known: 1) slope (gradient) and 2) point at which the line crosses the y-axis of the graph (intercept). The equation for straight lines is represented by:

$$Y = b_o + b_1 X_i + e_i$$

Where
Y = Outcome variable of interest (dependent variable)
β_0 = Y-intercept
β_1 = Slope
X_i = i^{th} subject's score on the predictor (independent variable)
ε_i = Residual or Error term $\left(\hat{X}_i - X_i\right)$

Slope:
Conceptually, the formula for slope is:

$$\frac{\text{cov}_{(x,y)}}{S_x}$$

Notice that the covariance is the numerator for both the slope and the correlation coefficients (page 20). Recall that the correlation coefficient provides a measure of how two random variables are associated in a sample. It is also a measure of the strength of the linear relationship between X and Y. Intuitively, the slope is an indicator of the relationship between two variables.

Therefore, if the slope is positive, then the correlation coefficient will be positive and vice versa. The correlation coefficient will always have the same sign as the slope.

Y-Intercept:
The formula for the Y-intercept is:

$$\hat{b}_o = \overline{Y} - \hat{b}_1 * \overline{X}$$

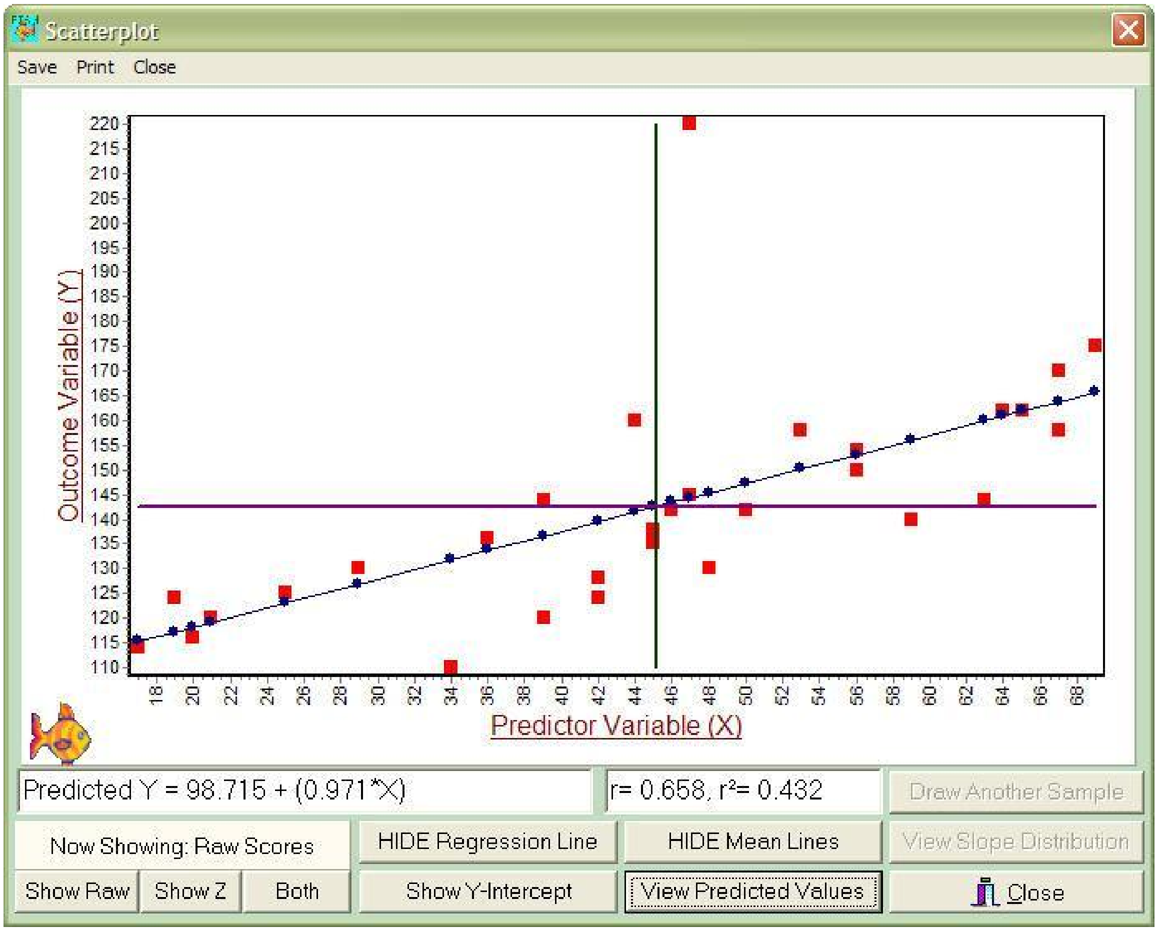
Regression Equation:
The least squares line is generally represented as:

$$\hat{Y} = \hat{b}_o + \hat{b}_1 * X$$

Note that FISH provides the regression equation under the "View Scatterplot" option.

Scatterplot Feature

It is important to examine a distribution of data graphically. Using the scatterplot option, FISH allows users view the distribution of the data.



By clicking "View Predicted Values", FISH will calculate the observed FISH also has the capability to calculate observed, predicted, and residual values, squared deviation scores, model sum of squares ($(\hat{Y} - \bar{Y})^2$), as well as calculations of the standard error of the estimate, standard error of the regression slope, and the 95% confidence interval for the regression slope.

Please refer to the User's Guide for a more thorough examination of FISH's graphics capabilities.

6 Using the Data Generator to Teach Statistical Concepts

FISH allows users to generate random data. Data is generated by defining a population and then sampling from that population. FISH's graphics capabilities complement the data generator and allow for abstract concepts such as sampling distributions to be concretely explained to students.

With the one variable analysis, the histogram feature is enhanced to allow the user to view the sampling distribution of the mean. From this screen, the user can generate additional samples in multiples of 1, 10, 100 or 10,000. These sample means are then added to the existing sampling distribution of the mean.

With the two variable analysis, the scatterplot feature is enhanced to allow users to view the sampling distribution of the regression coefficient. From this screen, the user can generate additional samples in multiples of 1, 10, 100 or 10,000. These sample regression coefficients are then added to the existing sampling distribution of the regression slope.

It is important to note that as each sample is drawn from the population, the data are entered into the spreadsheet on the main FISH page. It can be helpful to "toggle" back and forth from the graphics to show students that as samples are drawn, new statistics are calculated for each sample.

<u>Central Limit Theorem:</u>

The central limit theorem is an important, yet difficult concept for beginning (and sometimes intermediate!) statistics students. FISH's random data generator allows a sampling distribution for the mean to be created, and the program's graphics capabilities help to make the concept more tangible for students.

Using the one variable analysis, FISH creates a *sampling distribution of the mean.* A sampling distribution of the mean is a distribution of sample means when repeated samples are taken from a population. (It is important to note that sampling distributions can be created for any statistic, not just the mean.)

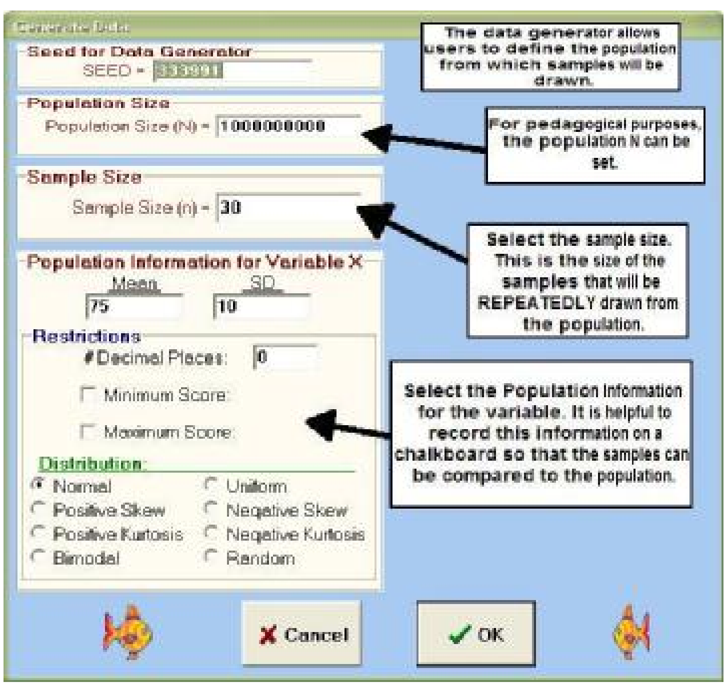
Recall that the *central limit theorem* is based on the characteristics of a sampling distribution when a reasonable large (N = 30) sample is drawn. The central limit theorem states that as sample size (N) gets infinitely large, the shape of the sampling distribution of the mean approaches the normal distribution with:

$$\text{mean} = \mu$$

$$\text{standard deviation} = \frac{s}{\sqrt{N}}$$

Recall that the standard deviation of a sampling distribution is called the standard error of the mean (see page 14). Therefore, as $N \to \infty$, the graph of the sampling distribution of the mean approaches normal with an expected value = μ and a standard error of $\frac{s}{\sqrt{N}}$. Because the sample size is in the denominator of the formula for standard error, as N increases, the standard error of the mean decreases. Intuitively, a large N will result in sample means that are closer in value to the population mean.

To simulate the central limit theorem using FISH, data must be generated using the data generator. See the User's Guide for the data generation procedure.

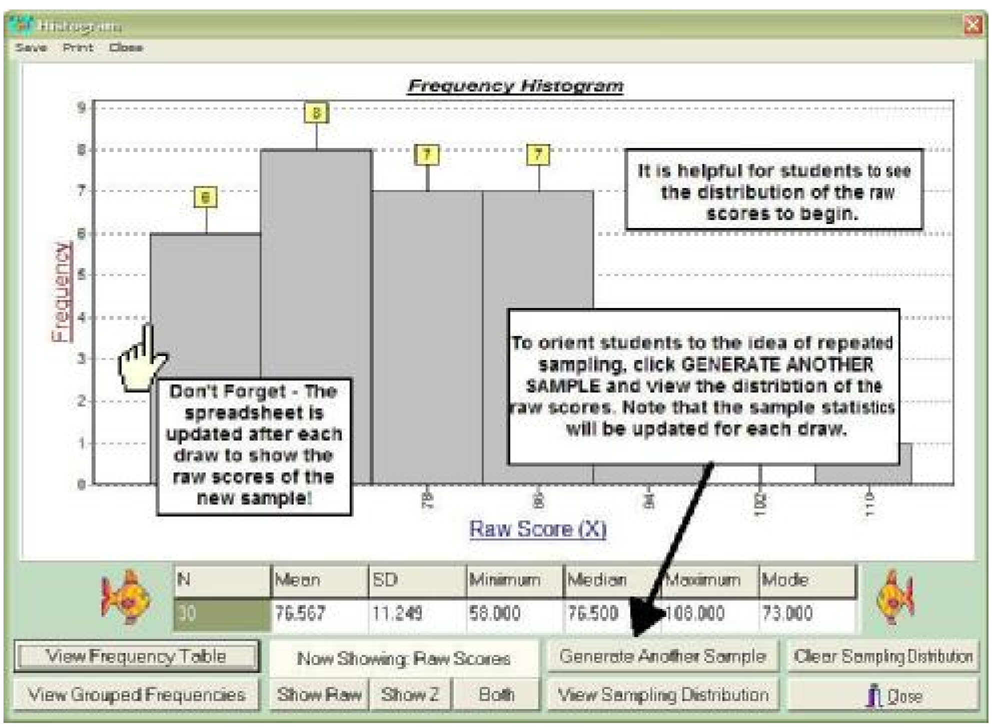


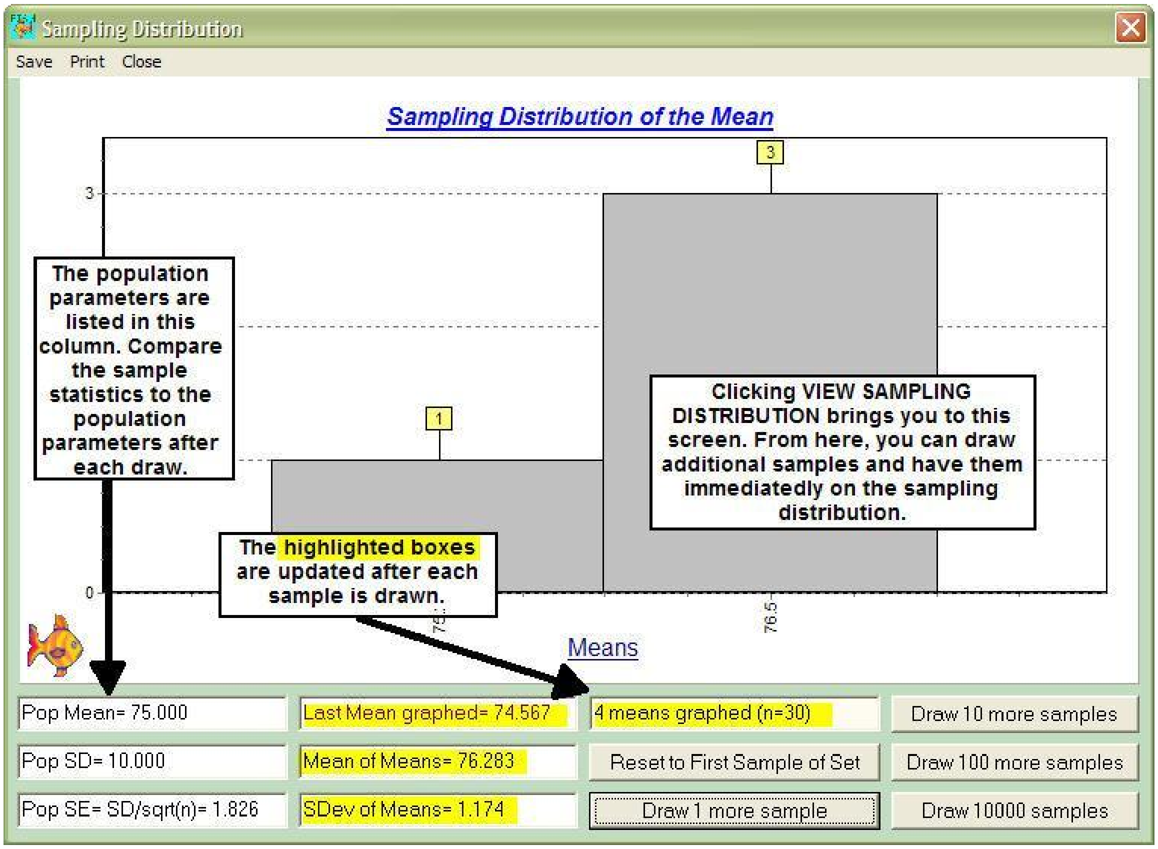
As you set the population parameters, it is helpful to explain to students that this is a contrived example. In reality, we rarely know the population parameters. If we knew the population parameters, we would not need to estimate them using statistics.

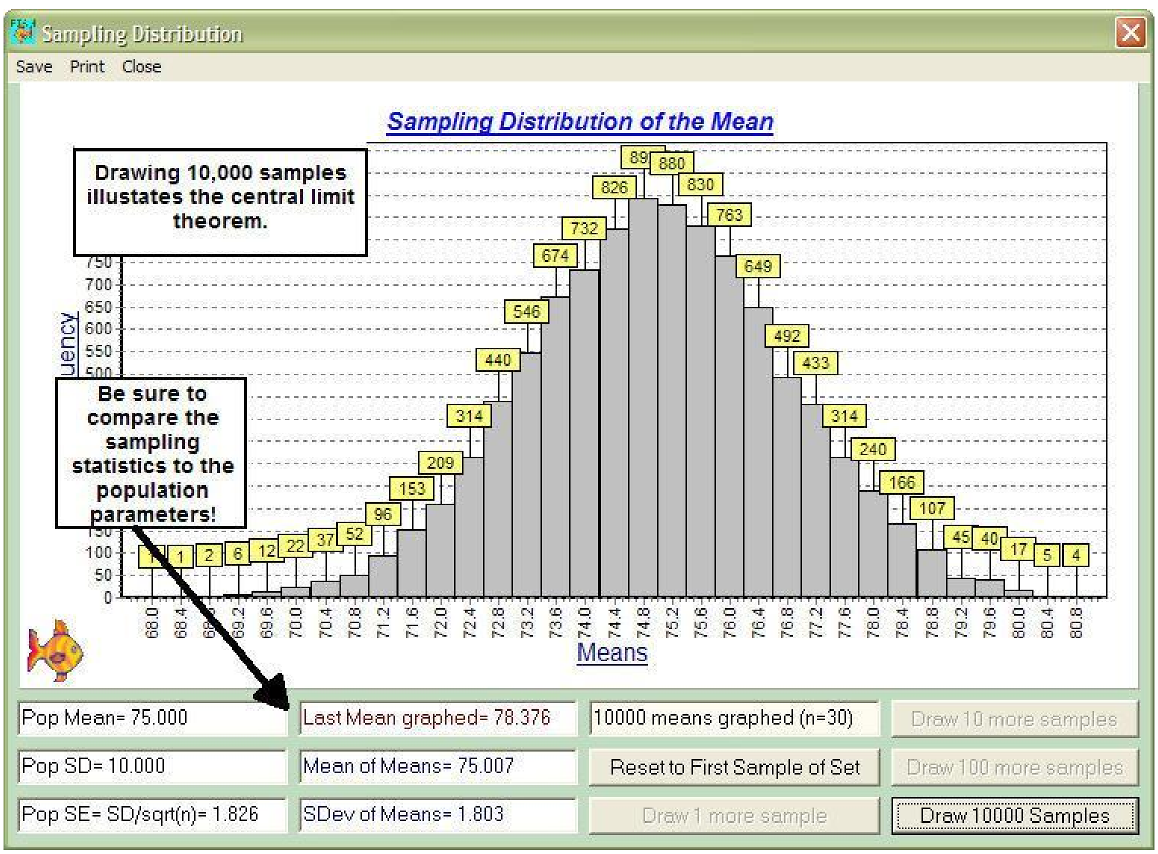
The goal of this learning activity is to concretely explain the central limit theorem. Therefore, it is helpful to create a scenario. An example would be: Ohio Stadium seats 105,000 people (the population N). We, as researchers, decide to conduct a survey of the people in the stadium to determine how much money the average Ohio State fan has in his/her wallet on game day. To complete our study, everyone in class is going to survey 30 people (the sample n). For some reason (and this NEVER happens, because it is a contrived example) Brutus the Buckeye mascot whispers in our ear “The population mean is \$75 (μ = 75) and the population standard deviation is \$10 (σ = 10).” Brutus then informs us that the data are normally distributed.

After students are given the contrived scenario, FISH can be used to create a sample (n = 30) for each member of the class. Students can record their sample means and standard deviations and watch as their mean is added to the distribution of sample means.

An example of how the activity would proceed follows:




Sampling Distribution
Save Print Close
Sampling Distribution of the Mean
The population parameters are listed in this column. Compare the sample statistics to the population parameters after each draw.
Clicking VIEW SAMPLING DISTRIBUTION brings you to this screen. From here, you can draw additional samples and have them immediatedly on the sampling distribution.
The highlighted boxes are updated after each sample is drawn.
Means
Pop Mean= 75.000
Last Mean graphed= 74.567
4 means graphed (n=30)
Draw 10 more samples
Pop SD= 10.000
Mean of Means= 76.283
Reset to First Sample of Set
Draw 100 more samples
Pop SE= SD/sqrt(n)= 1.826
SDev of Means= 1.174
Draw 1 more sample
Draw 10000 samples


Sampling Distribution
Save Print Close
Sampling Distribution of the Mean
Drawing 10,000 samples illustates the central limit theorem.
Be sure to compare the sampling statistics to the population parameters!
Means
Pop Mean= 75.000
Last Mean graphed= 78.376
10000 means graphed (n=30)
Draw 10 more samples
Pop SD= 10.000
Mean of Means= 75.007
Reset to First Sample of Set
Draw 100 more samples
Pop SE= SD/sqrt(n)= 1.826
SDev of Means= 1.803
Draw 1 more sample
Draw 10000 Samples

At this point, it should be pointed out that our population of 105,000 in the example is not large enough to draw 10,000 samples of size 30. But since this is a *contrived* example, students typically get the picture.

The next step in the learning activity is to explore the central limit theorem further. Notice that the central limit theorem did NOT say that the population has to be distributed normally. To confirm this, repeat the exercise using a non-normally distributed population. Further, you could highlight that as long as the sample size is reasonably large ($n > 30$) the distribution of sample means will be approximately normal. This can be confirmed by adjusting the sample size. Using FISH, students can clearly see that when n is reasonably large, a distribution of sample means will approximate the normal distribution even if the means come from a population that is non-normal.

It should be noted that the central limit theorem applies to all populations and can be used to estimate the mean from any population when reasonably large samples are drawn. The central limit theorem was named such because the idea is CENTRAL to statistics. The central limit theorem allows us to assess how likely it is that a random sample will yield a value for the sample mean ($\overline{X}$) that is any given distance from the population mean (μ). This concept will be used throughout any statistics course.

<u>Distribution of the Slope</u>

COMING SOON – WATCH THE WEB FOR UPDATES!
http://oak.cats.ohiou.edu/~brooksg/fish.htm

7 References

The following statistics references were consulted in the writing of this manual:

Field, A. (2000). *Discovering statistics using SPSS for windows*. Thousand Oaks, CA: Sage.

Harris, M. B. (1998). *Basic statistics for behavioral science research*(2nd ed.). Needham Heights, MA: Allyn & Bacon.

Howell, D.C. (2004). *Fundamental statistics for the behavioral sciences*(5th ed.). Belmont, CA: Brooks/Cole – Thompson Learning.

Lemeshow, S. (2003, June). Applied Regression Analysis. In *2003 Summer Program in Applied Statistical Methods*, conducted at The Ohio State University, Columbus, OH.

8 User Agreement

Use of the FISH software program provided with this Agreement constitutes acceptance of these terms and conditions of use. If you do not agree to the terms of this agreement, do not use the FISH software program.

LICENSE

FISH is a copyrighted program and is NOT public domain. The user is granted license, not ownership, to use the FISH software program on any computer subject to the restrictions described in the User Agreement and Disclaimer.

FISH is Freeware. The user is licensed to make an unlimited number of exact copies of the FISH software program, to give these exact copies to any other person for their personal use, and to distribute the FISH software program in its unmodified form only via disk, email, or local area network. If these methods of distribution are unavailable, any person wanting to use the FISH software program should be directed either to contact the author or to visit the author's internet web site (the URL is provided below and may be posted on any web site).

If you find the program useful, if you copy it for others, if you find problems or bugs in the program, or if you use the program for teaching, educational, or consulting purposes, you are requested to inform the authors:

Author:	Gordon P. Brooks, Ph.D.
Address:	McCracken Hall, Ohio University, Athens, OH 45701
Telephone:	740-593-0880
Fax:	740-593-0477
Email:	brooksg@ohiou.edu
Web Page:	http://oak.cats.ohiou.edu/~brooksg/fish.htm

TERMS OF USE

The FISH software program may be used and copied for personal use subject to the following The FISH software program may be used and copied for personal use subject to the following license restrictions:

- the FISH software program shall be copied and supplied in its original, unmodified form;
- the FISH software program shall not be sold or used for profit, nor may any amount or fee be charged for use, rental, lease, or distribution of the program;
- the FISH software program shall not be included or bundled with any other goods or services;
- the FISH software program may not be decompiled, disassembled, or otherwise modified; any such unauthorized use without expressed, written permission granted by the author shall result in immediate and automatic termination of this license.

DISCLAIMER OF WARRANTY

Every effort has been made to ensure the accuracy of the FISH program, the algorithms and subroutines used, and the results produced by the FISH software program, both on screen and printed.

However, no warranty is expressed or implied concerning the function or fitness of the FISH software program, subroutines, or results provided by the program. That is, the FISH software program is provided on an "as is" basis without warranty of any kind. The author shall have neither liability nor responsibility to any person or entity with respect to any liability, loss, or damage directly or indirectly arising from the use of or inability to use the FISH software program or the results of the analyses provided by the FISH software program, even if the author has been advised of the possibility of such damages or claims. In no event shall any liability exceed the license fee paid to the author of the FISH software program. In the event of invalidity of any provision of this license, the user agrees that such invalidity shall not affect the validity of the remaining portions of this license.

User's Guide and Instructor's Manual Prepared by Holly Raffle.