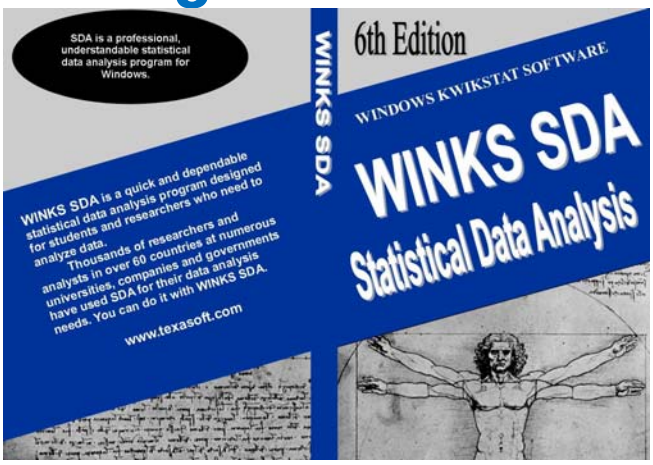
WINKS SDA

Windows KwikStat
Statistical Data Analysis

Getting Started Guide



Sixth Edition

Do these tutorials first

This series of tutorials provides a quick start to using WINKS. Feel free to duplicate this Getting Started Guide to use in your classroom, business, or for your personal use.

Contents

Note: This document is online at
http://www.texasoft.com/college.html
(Revision: June 2010)

Getting Started Tutorials

- An Overview of WINKS SDA
- Installation
- Tutorial 1 – Analyzing Summary Data
- Tutorial 2 – Entering Data to Analyze
- Tutorial 3 – Using Data from an Existing File
- Tutorial 4 – Using SDA Graphs
- Tutorial 5 – Importing data from Excel

★ **Name Change? –** If you've used WINKS before, you'll notice that we now refer to the program as WINKS SDA or just SDA – since Microsoft also has a program called WINKS, we think it will make it easier to refer to WINKS as SDA, which stands for Statistical Data Analysis.

★ **Reference -**When you report SDA results in a professional publication, use the following citation: **TexaSoft, WINKS SDA Software Ver. 6, Cedar Hill, Texas, 2007.**

An Overview of SDA

Statistical data analysis allows you to make informed decisions. WINKS SDA is a statistical data analysis program designed by professional statistical consultants and researchers that allows you to calculate commonly used statistical data analysis procedures and graphs. A few of the main benefits of SDA include:

1. The native file format for WINKS SDA is an ".sda" file. SDA can also read .DBF, Excel (.XLS) and text files (.CVS, .TXT, .DAT etc.).
2. Many SDA procedures report information that allows you to make a reasonable conclusion based on the statistical findings -- unlike other statistical programs that make you run each part of an analysis separately and piece your answers together. For example, in SDA's Analysis of Variance procedures, multiple comparisons are automatically performed and boxplots are created to show you where specific pair-wise differences lie (unlike most other statistical programs.)
3. SDA often provides warnings when common assumptions are not met and makes suggestions about how a particular outcome could be interpreted. For many tests, an explicit hypothesis is stated to help you understand the exact intention of the analysis.

Installation

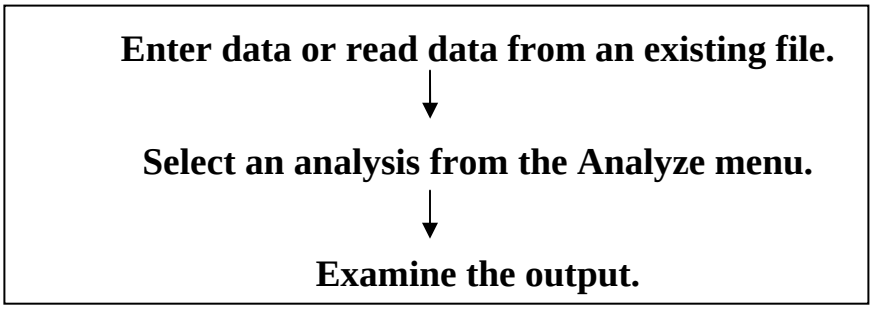
Place the WINKS SDA distribution CD in your CD Drive. This should automatically start the SDA Installation procedure. If the installation does not automatically begin: Click START/Run and enter D:\SETUP.EXE (Use your CD drive name.) Or, to make sure you have the latest release, you can download the installation program at

http://www.texasoft.com/updates

Follow the installation instructions on the screen. When you begin SDA for the first time, a dialog box will appear, asking for a key code. Enter the code provided with your CD. (Some vendors provide the code as a "Serial Number.") After the installation you will be prompted to choose program options. Select options you want to change, or leave the default options in place (recommended) and choose Ok. You can change these options at any time by choosing "Change Setup Options" from the Help menu.

The Analysis Process

Here is a brief look at the data analysis process:



Tutorial 1 – Analyzing Summary Data

Suppose you have the following (summarized) information about two groups you wish to compare:

Data for Group 1

Mean = 23.44, Standard Deviation = 3.41, N = 8

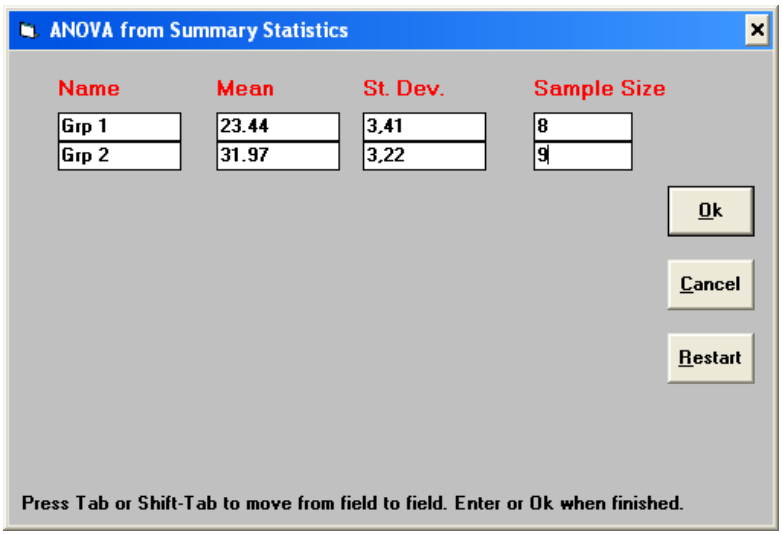
Data for Group 2

Mean = 31.97, Standard Deviation = 3.22, N = 9

To perform an Independent Group Student's t-test from summarized data, follow these steps:

Step 1: Begin SDA. Select **Analyze/t-tests and ANOVA/ Ind. Group from Summary Data.**

Step 2: When prompted to enter the number of groups, **enter 2**. This dialog box appears:



Step 3: Enter the numbers as shown in the dialog box and **Click Ok**. The following output (abbreviated) appears.

```
------------------------------------------------------------------------
Independent Group Analysis                                 Summary Data
------------------------------------------------------------------------

Group Means and Standard Deviations
-----------------------------------
Grp 1:  mean = 23.44          s.d. = 3.41            n =  8
Grp 2:  mean = 31.97          s.d. = 3.22            n =  9


Mean Difference = -8.53     S.E. Difference = 1.61417

Test for Equality of Variance
-----------------------------
This preliminary test determines which version of the t-test to perform.
Test equality of variance: F = 1.12 with (7, 8) D.F.  p = 0.433 (two-tail)

Note: Since the p-value for equality of variance is greater than 0.05,
use the Equal variance t-test results.

Independent Group t-test Hypotheses
-----------------------------------
Ho: There is no difference between means.
Ha: The means are different.

Independent Group t-test on Summary Data
------------------------------------------------------------------------
  Equal variance: Calculated t= -5.3 with  15  D.F.  p <= 0.001 (two-tail)
Unequal variance: Calculated t= -5.28 with 14.5 D.F. p <= 0.001 (two-tail)

(For a one-sided test, you must adjust the p-value according to
the direction of your alternative hypothesis.)

Confidence Interval
-------------------
A 95% Confidence Interval about the mean difference is: (   -11.9715 to -5.0885)
```

This tests the assumption that variances are equal.

WINKS often describes the specific hypothesis tested.

Report the appropriate t-test based on whether variances are assumed equal or not.

```
CI is based on a standard error of 1.6142 and a tabled t-value of 2.132 with 15 d.f.

 Effect Size Measures
 --------------------
 Cohen's d = -2.743, a small effect size.    (Ref:Rosenthal and Rosnow, 1991)
 Hedge's g = -2.573 (an adjustment of Cohen's d for sample sizes <10.)

 Possible Writeup
 ----------------
 The means of the two groups were significantly different.
 t(15) = -5.3, p <= 0.001

 Statistics were performed using WINKS SDA Software (Texasoft, Cedar Hill, TX.)
 Statistical decisions were made at p=0.05 unless otherwise stated.

 For these data, the Mean(SD) of DATA for GROUP = Grp 1 is 23.44(3.4100), N= 8,
 and the Mean(SD) of DATA for GROUP = Grp 2 is 31.97(3.2200), N= 9.

 (Results must be interpreted in the context of the practical {i.e. clinical}
 implications of any observed differences.)
```

WINKS often provides help in writing up your results.

The “test for equal variance” indicates which version of the t-test to use. Since this test is non-significant (p=0.433) you should use the “Equal Variances t-test” (unless you have some other reason to use the more conservative Unequal variances test.) Either way, your conclusion is to reject the null hypothesis (Ho) since p <.001 which show evidence for the alternative that the means are different.

Other Summarized Data Analyses

Other analyses in SDA that you can run from summary data include:

- Independent Group Means Comparison – Tests and Analysis of Variance
- Single Sample t-test
- Dunnett’s Test
- Contingency Table – Chi-Square Analysis -- also including Yate’s Test, Fisher’s Exact Test, Relative Risk, Odds Ratio, Sensitivity and Specificity, and more.
- Goodness-of-Fit Analysis
- Test of a Difference between Two Proportions
- P-value determination
- Bar and Pie Charts
- Kappa – Interrater reliability (Professional Edition)
- Pie and Bar Charts

Tutorial 2 – Entering Data to Analyze – Comparing Means

This example shows how to enter “raw” data into SDA and use that information to perform an analysis and display a graph. Suppose you are testing the effectiveness of four kinds of hog feed (to gain weight). You randomly assign 15 hogs to 4 groups, feed each group one of the feeds for

a month, and observe weight gain. You want to know which type of feed produced the most (average) weight gain for the four groups.

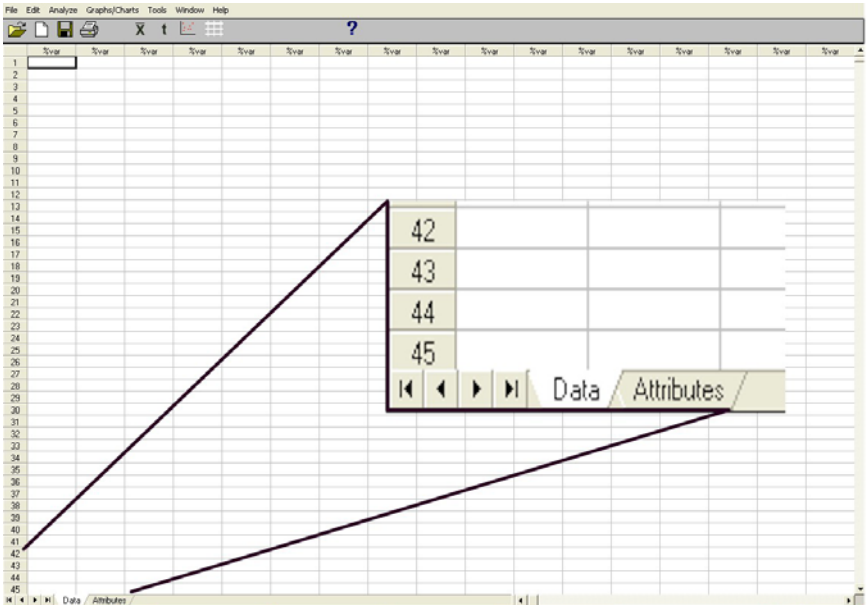
GROUP(Feed)	*OBS(WEIGHT)*
1	60.8
1	67.0
1	54.6
1	61.7
2	78.7
2	77.7
2	76.3
2	79.8
3	92.6
3	84.1
3	90.5
4	86.9
4	82.2
4	83.7
4	90.3

Note: When you create a data set where data are in groups, you should include a variable that indicates which observation (subject) belongs to which group. (in this case the variable will be named FEED.) You can select numbers for the group names (such as 1,2,3,4) or letters (A,B,C,D) or even names (ACME, DELUX, STANDARD, BESTFEED). This group name is what WINKS uses to identify which observation belongs to which group.

To perform this analysis:

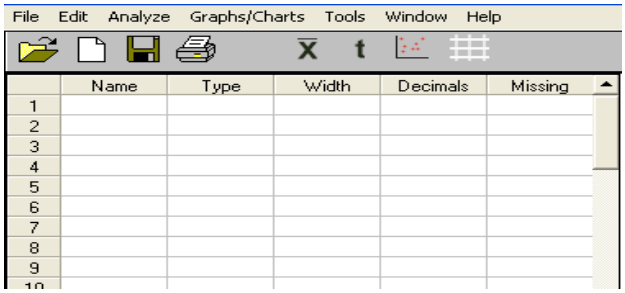
1. Enter the data into a SDA database
2. Select the type of analysis to perform
3. Observe and interpret the output

Step 1: When you begin SDA a blank data grid appears as shown below:



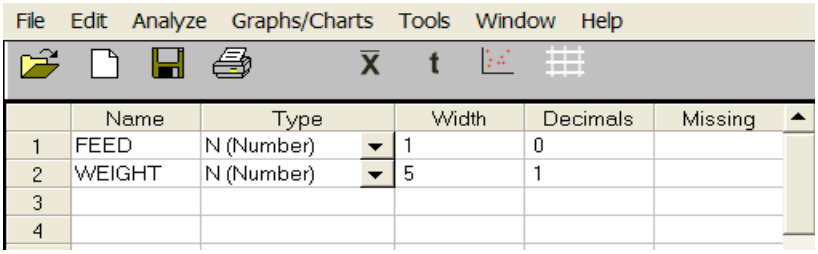
You could immediately enter data, but we recommend that you follow this procedure:

Step 2: Click on the Attributes tab at the bottom of the SDA grid to display the following screen:



Define your data set variables in this grid. For this case you need a grouping variable (FEED) and an observation variable (WEIGHT).

- In the Name column enter *FEED* in row 1 and *WEIGHT* in row 2
- In the Type column select (N) Number for both variables.
- In the Decimals column enter 0 for *Feed* and 1 for *Weight*
- In the Width column enter 1 for *Feed* and 5 for *Weight*
- Leave the Missing column blank for now.
- See the completed information below:



Step 3: Click on the Data tab to return to the data grid. Enter the data from the data list above as shown below:

File Edit Analyze Graphs/Charts Tools Window Help

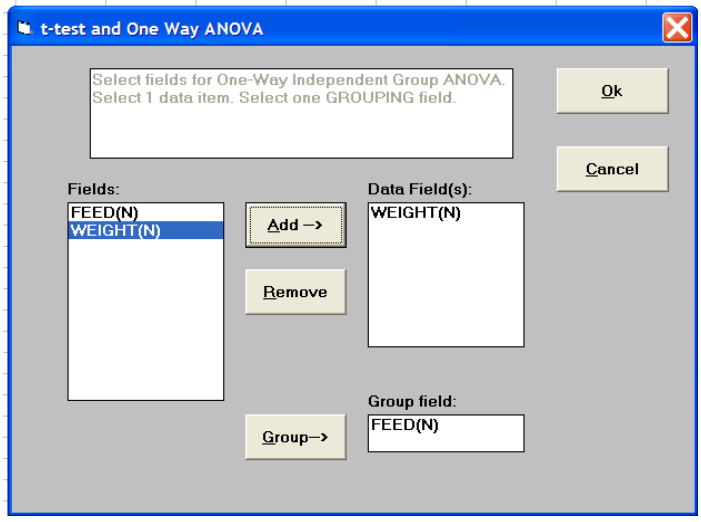
	FEED	WEIGHT	%var	%var	%var
1	1	60.8			
2	1	67.0			
3	1	54.6			
4	1	61.7			
5	2	78.7			
6	2	77.7			
7	2	76.3			
8	2	79.8			
9	3	92.6			
10	3	84.1			
11	3	90.5			
12	4	86.9			
13	4	82.2			
14	4	83.7			
15	4	90.3			

Notice that for this example 1, 2, 3, and 4 represent *Feed* brands – you could use A, B, C, D or any alpha-numeric name for the group names. The *Weight* field contains the variable you observed. In this case it is weight gain.

Step 4: When you have finished entering the data, **select File/Save As** and save the data set with the name Hog Tutorial. (You don't have to save the file to do an analysis, but we recommend it.)

Step 5: Select **Analyze/t-test and ANOVA/ Independent Group (t-test/ANOVA)**.

Step 6: Choose what fields to use. A dialog box appears allowing you to select which fields to use for this analysis. **Select the *FEED*** field, and click on the **Group** button. Then **select the *WEIGHT* field** and click on the **Add** button. Your field choices will look like the dialog box in the figure below. **Click Ok** to continue.



Step 7: The output (abbreviated below) compares the 4 means:

```
------------------------------------------------------------------------------
Independent Group Analysis                    C:\AWINKSRC\Hog Tutorial.SDA
------------------------------------------------------------------------------

Grouping variable is FEED
Analysis variable is WEIGHT

Group Means and Standard Deviations
-----------------------------------
1:  mean = 61.025             s.d. = 5.0822           n =  4
2:  mean = 78.125             s.d. = 1.4886           n =  4
3:  mean = 89.0667            s.d. = 4.4276           n =  3
4:  mean = 85.775             s.d. = 3.5976           n =  4


Analysis of Variance Table

Source              S.S.        DF             MS      F         Appx P
------------------------------------------------------------------------------
Total            1923.41        14
  Treatment      1761.24         3         587.08   39.82          <.001
  Error           162.17        11          14.74

Error term used for comparisons = 14.74 with 11 d.f.

                                                                  Critical q
Newman-Keuls Multiple Comp.        Difference   P     Q             (.05)
-------------------------------------------------------------------------
    Mean(3)-Mean(1) =                28.0417   4    13.523        4.256 *
    Mean(3)-Mean(2) =                10.9417   3     5.277         3.82 *
    Mean(3)-Mean(4) =                 3.2917   2     1.587        3.113
    Mean(4)-Mean(1) =                  24.75   3    12.892         3.82 *
    Mean(4)-Mean(2) =                   7.65   2     3.985        3.113 *
    Mean(2)-Mean(1) =                   17.1   2     8.907        3.113 *

Homogeneous Populations, groups ranked

               Gp Gp Gp Gp
                1  2  4  3
                     ------
               ---
                  ---

This is a graphical representation of the Newman-Keuls multiple comparisons
test. At the 0.05 significance level, the means of any two groups
underscored by the same line are not significantly differe

Simultaneous 95% Confidence Limits
----------------------------------
Significant comparisons based on Conf. Limits indicated by ***.
CI uses Tukey-Kramer procedure. P-values reflect a Bonferroni adj      ent.
Error term used = 14.74 with 11 d.f.

    Group                                              Simultaneous 95%
    Comparison           Difference   p-value       Confidence Limits
-------------------------------------------------------------------------
    Mean(3)-Mean(1) =       28.0417     <.001    (19.2161, 36.8672)  ***
```

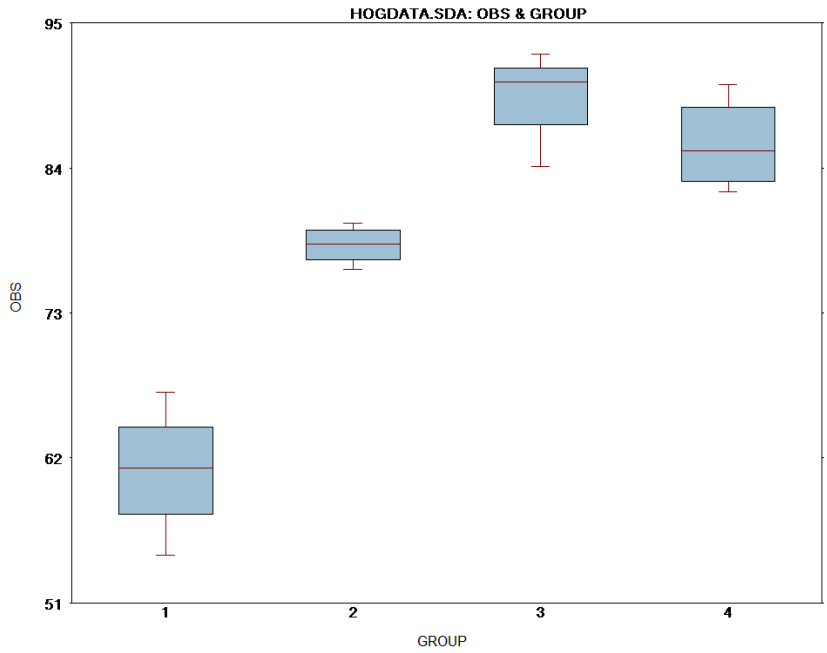
The ANOVA table tests if any of the means are not equal. In this case p <.001 indicates there is a difference in some means.

The multiple comparison test indicates which means are different at the alpha=0.05 significance level (those marked with an *.) Note: You can select which multiple comparison procedure you use by clicking on the Option button when selecting variables for the analysis.

Another way to assess significance is with confidence limits. These are selected by clicking on the Options button when selecting variables.

```
    Mean(3)-Mean(2) =        10.9417     0.020    (3.0214, 18.8619)   ***
    Mean(3)-Mean(4) =        3.2917      1.000    (-3.1628, 9.7462)
    Mean(4)-Mean(1) =        24.75       <.001    (17.4173, 32.0827)  ***
    Mean(4)-Mean(2) =        7.65        0.100    (1.6743, 13.6257)   ***
    Mean(2)-Mean(1) =        17.1        <.001    (11.1243, 23.0757)  ***

Note: Because different multiple comparisons procedures are based on different
methods, they may not completely agree for marginally significant comparisons.

Possible Writeup
----------------

A one-way Analysis of Variance was performed to test the hypothesis
that the means across categories of FEED were equal.
In the presence of significance for the omnibus ANOVA test, a
Newman-Keuls multiple comparison test is used to perform pairwise comparisons.
Statistics were performed using WINKS SDA Software (Texasoft, Cedar Hill, TX.)
Statistical decisions were made at p=0.05 unless otherwise stated.

The means across categories of FEED were found to be different
F( 3, 11)=39.82, p <.001.
A Newman-Keuls multiple comparison procedure was performed at the alpha=0.05
significance level to determine specific pairwise differences.

For these data,
the Mean(SD) of WEIGHT for FEED = 1 is 61.025(5.0822), N= 4,
the Mean(SD) of WEIGHT for FEED = 2 is 78.125(1.4886), N= 4,
the Mean(SD) of WEIGHT for FEED = 3 is 89.0667(4.4276), N= 3,
and the Mean(SD) of WEIGHT for FEED = 4 is 85.775(3.5976), N= 4.

(Results must be interpreted in the context of the practical {i.e. clinical}
implications of any observed differences.)
```

Step 8: To display a graphical interpretation of the analysis, select **Graphs/Display Graph**. The default graph shows a box and whiskers plot comparison in the following graph:

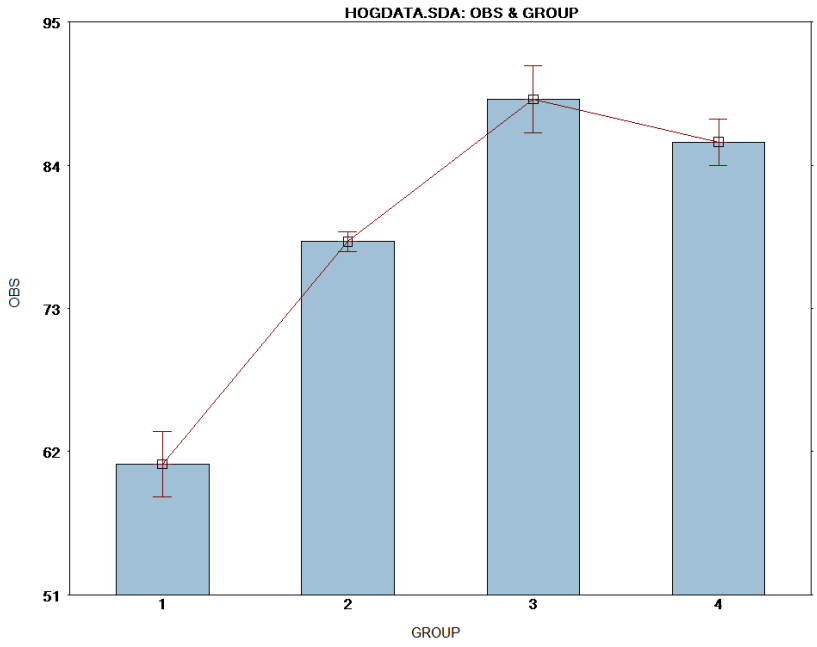


Step 9: From the group comparison plot, you can select a variety of display options**. Select Edit/Options** to display the options menu for the graph.

Step 10: For this example, check the following options (and uncheck all other options) on Graph Options dialog box:

- Display Means
- Error Bar (+/- St. Error of the Mean)
- Connect Means
- Bar Graph of Means

The following graph is displayed. Print the graph by selecting File/Print or copy the file to the Windows clipboard by selecting Edit/Copy to Clipboard.

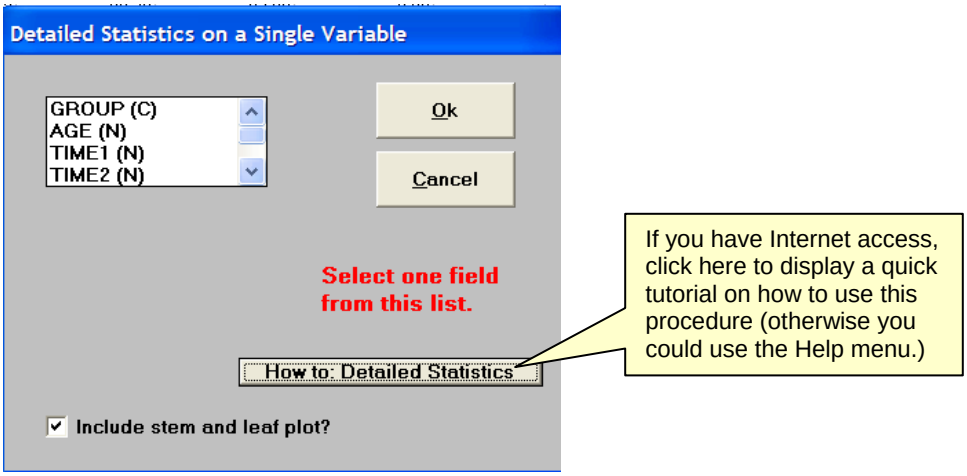


Tutorial 3 – Using Existing Data

This example shows you how to open an existing data set, display descriptive statistics on a single variable, and display a histogram. Follow these steps:

Step 1: **Select the file named EXAMPLE.SDA**. Once the file opens you will see the data in the data grid.

Step 2: To perform a Descriptive Statistics analysis on this data select **Analyze, Descriptives, Detail/One Variable**. A dialog box appears where you can choose a field name**. First click on the checkbox labeled "include stem and leaf plot?" Click on *AGE*, and Ok**, as shown in the following dialog box.



Step 3: A dialog box is displayed allowing you to specify options for the Stem and Leaf plot. **Click on the checkbox labeled "Split stem leaf value in half."** Click **Ok.**

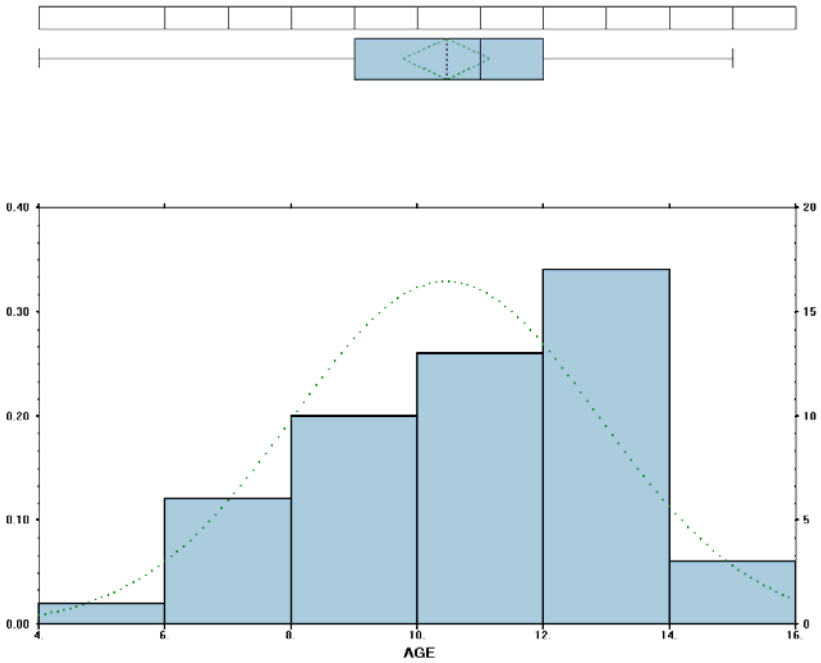
Step 4: The output contains descriptive statistics from the data in the *AGE* field. Scroll this window to view all of the output.

```
------------------------------------------------------------------------------
Descriptive Statistics                              C:\AWINKSRC\EXAMPLE.SDA
------------------------------------------------------------------------------
Variable Name is AGE

N         = 50                    Missing or Deleted = 0
Mean      = 10.46                      St. Dev (n-1) = 2.42613
Median    = 11.00                        St. Dev (n) = 2.40175
Minimum   = 4.00                              S.E.M. = 0.34311
Maximum   = 15.00                          Variance  = 5.88612
Sum       = 523.00                        Coef. Var. = 0.23194
Skewness  = -.512                         Kurtosis   = -.261
------------------------------------------------------------------------------
Percentiles:                                 Tukey Five Number Summary:
0.0%          = 4.00    Minimum              Minimum  = 4.00
0.5%          = 4.00                         25th     = 9.00
2.5%          = 4.55                         Median   = 11.00
10.0%         = 7.00                         75th     = 12.00
25.0%         = 9.00    Quartile             Maximum  = 15.00
50.0%         = 11.00   Median
75.0%         = 12.00   Quartile
90.0%         = 13.00
97.5%         = 14.725
99.5%         = 15.00                        Test for normality results:
100.0%        = 15.00   Maximum              D = .148     p = 0.008

Five number summary consists of the 0, 25, 50, 75 and 100th percentiles.

Confidence Intervals about the mean:
------------------------------------------------------------------------------
80 % C.I. based on a t(49) critical value of 1.3 is (10.01396, 10.90604)
90 % C.I. based on a t(49) critical value of 1.68 is (9.88358, 11.03642)
95 % C.I. based on a t(49) critical value of 2.01 is (9.77035, 11.14965)
98 % C.I. based on a t(49) critical value of 2.41 is (9.63311, 11.28689)
```

```
99 % C.I. based on a t(49) critical value of 2.68 is (9.54047, 11.37953)

Extreme values (Case Number)
-----------------------------------
 LOWEST 5         |  HIGHEST 5
-----------------------------------
 4.00(20)            15.00(36)
 6.00(17)            14.00(12)
 6.00(24)            14.00(38)
 7.00(9)             13.00(16)
 7.00(46)            13.00(27)


The normality test suggests that the data are not normally distributed.
The test for normality is a modified Kolmogorov-Smirnov test based on
papers by Lilliefors and Dallal & Wilkinson.
Some statisticians recommend that this test is only valid if n is less than 50.

Stem and Leaf Display

Field=AGE, Stem unit= 1 , N =  50

Frequency    Each leaf represents 1  observations.

 17        0|46677778888999999
(33)       1|000001111111122222222233333333445

The numbers to the far left are the counts of leafs
for a particular branch. The number with parentheses ()
indicates which branch contains the median of the data.
```

*Note: The "Tukey Five Number Summary" output will differ slightly if the "Display outliers" option was selected in the Setup options (**Help/Change Setup Options**.)*

Step 4: To display a graph of the data select **Graph/Display Graph.** A screen appears showing a histogram of the AGE data. **Select Edit/Display Normal Curve** to add a bell curve (normal curve) to the graph. The graph is shown below:



Step 5: To combine the graphic and text output, first select **Edit/Copy to clipboard** while viewing the graph. Then select **File/Exit** to return to the text output. Place your cursor at the end of the output and select **Edit/Paste**. The graph is pasted at the end of the output file. **Size the graph** by clicking on it. Sizing buttons appear at each corner. Click and drag on one of the sizing buttons to make the graph smaller.

Step 6: You can now print the output including the graph, or **select Edit/Select** all and **Edit/Copy** then paste the output into a word processor.

Tutorial 4 – Using SDA GRAPHS

SDA includes two types of graphs:

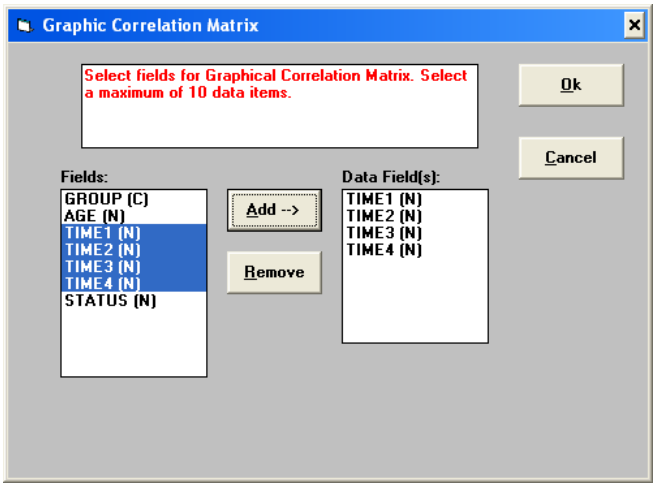
- **Analytic graphs** – are designed to help you understand the results of statistical tests.
- **Excel graphs** –displayed in Microsoft Excel (you must have Excel installed on your computer for this option to work.) (You can edit the resulting graph in Excel.)

Graph Example 1: This **Analytic graph** allows you to examine the relationship between several pairs of variables at once.

Step 1: Open the data set named EXAMPLE.SDA.

Step 2: Select **Graphs/Charts** and **Correlation Matrix Scatterplots** from the menu.

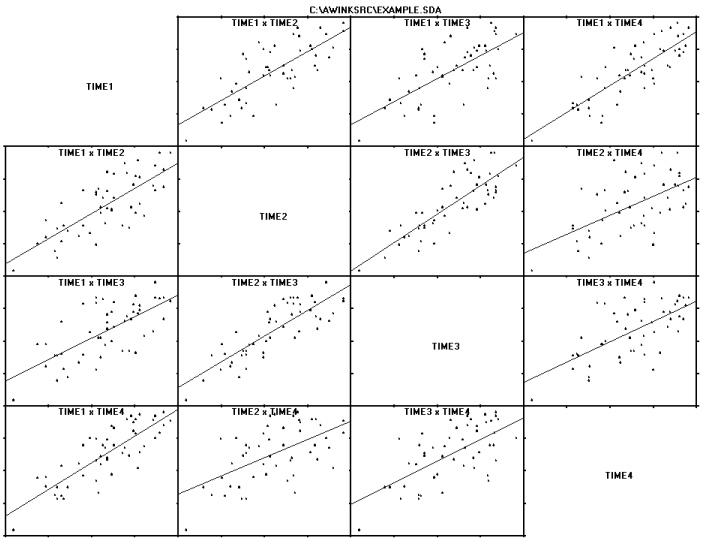
Step 3: From the variable selection dialog box, select the variables TIME1 to TIME4 as shown below by highlighting the variables names and click **Add.**



Step 4: Click **Ok** and the matrix of graphs will be created, as shown below.

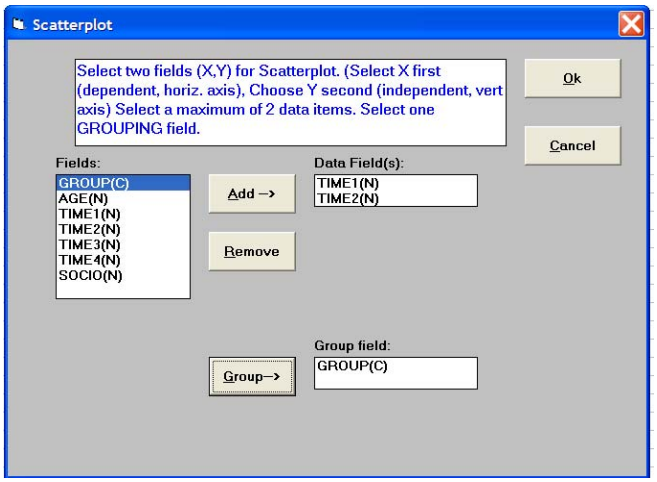
Step 5: Select **Edit/Display Regression Lines** to display regression lines on each graph.

This graph allows you to examine the relationships of several variables at once to discover which may be highly correlated and which are not.

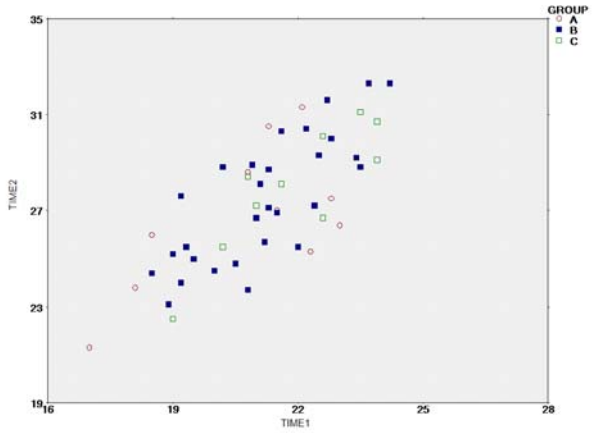


Graph Example 1a:

Step 1: With the same data set opened (Example) select Graphs/Charts and X-Y Scatterplot, then select Time1 and Time2 as the data fields and Group as the Group field.



Click OK to display the following graph:



This graph shows the relationship of Time1 to Time2, with different colors and symbols for each group category.

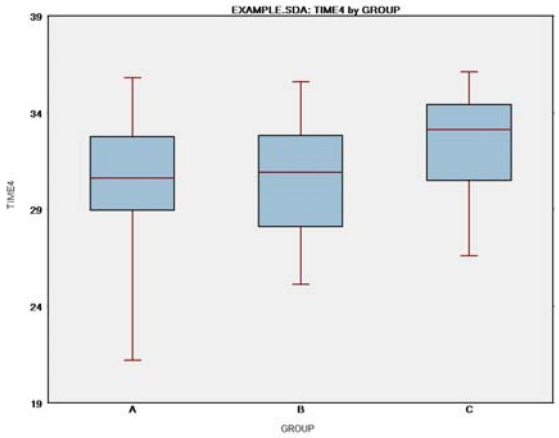
Step 2: Select Edit/Options. A dialog box appears. Select Regression Line and click OK. Three regression lines (one for each group) are displayed.

Step 3: Move your cursor to any point displayed on the graph and click. Information about that point is displayed on the graph. This is a good way to identify specific points of interest. Select File/Exit to return to the data screen.

Graph Example 1b: (Box and Whiskers Plot)

Step 1: Using the same data set, select Graph/Charts / By Group Plots (Boxplots, etc.)

Step 2: Select Time4 as the Data field and Group as the Group Field. Click Okay to display the comparative box plots



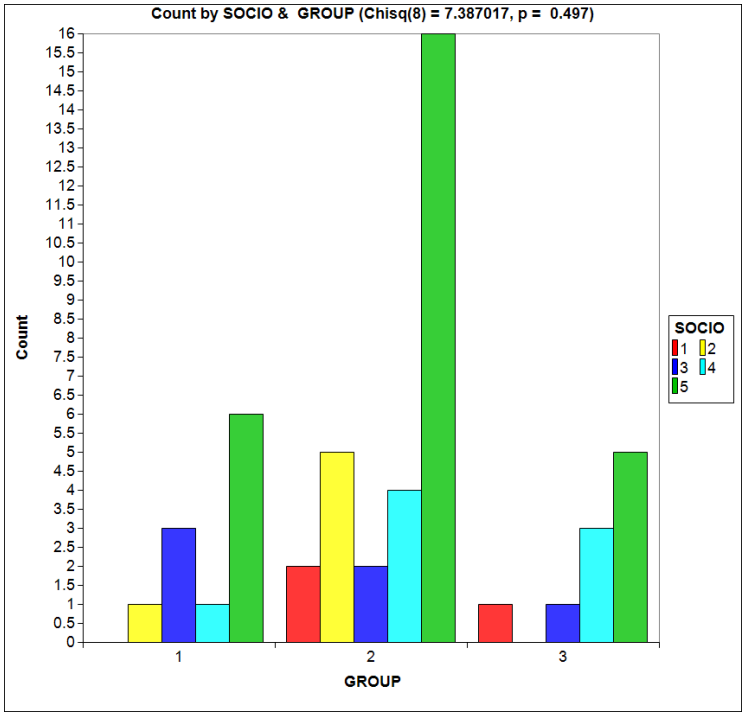
Step 3: Right click anywhere on the plot to display plot options. Select change box color, select red and OK and OK again to display the plot with your selected color. Experiment with other options from the same menu.

Graph Example 1c: (Grouped Bar Chart from Frequencies)

Step 1: Open the data set named EXAMPLE.SDA (you should already have it opened.)

Step 2: Select Graphs/Charts – Bar Charts - Group Bar Charts (from frequencies)

Step 3: Select SOCIO as the Data Field and GROUP as the Group field, Click Ok and the following chart appears.



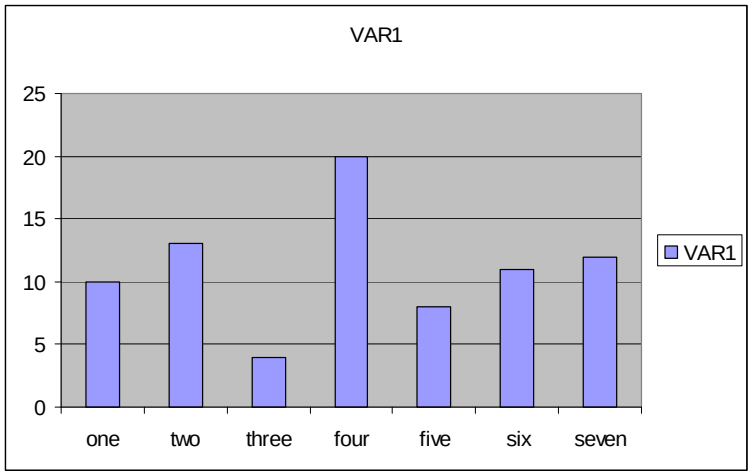
Step 4: Choose Edit/ Options or right click anywhere outside the graph to display options. On the Color options tab, click on Pastel. Then click on the Labels/Ok tab and click OK to display the graph using pastel colors.

Graph Example 2: This example shows how to create a bar chart in Excel.

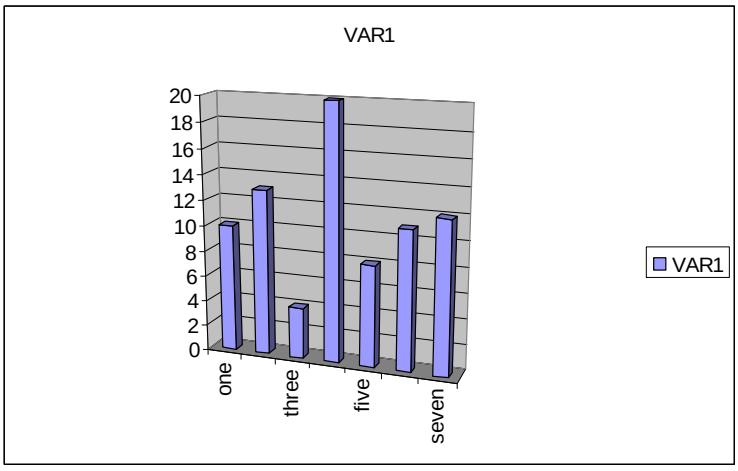
Step 1: Open the data set BARCHART.SDA (File/Open dataset)

Step 2: Select **Graphs/Bar Charts / Excel Bar Chart (from summarized data**). *(Be sure to select the Excel version of the Bar Chart.)* If you do not have Excel loaded on your computer, or if it is an incompatible version of Excel, you will get an error message when you attempt to create an Excel chart.

Step 3: Click on ***VAR1*** in the field box and **Add**. Click on ***LABEL*** variable name in the Field box and click on the **Label** button. Click **Ok**. The Excel program opens and within the worksheet a bar chart is displayed as shown below.



Step 4: Right click on graph and Select Chart Type. Select the 3-D chart (middle left option) to display the graph in a 3D format as shown below:



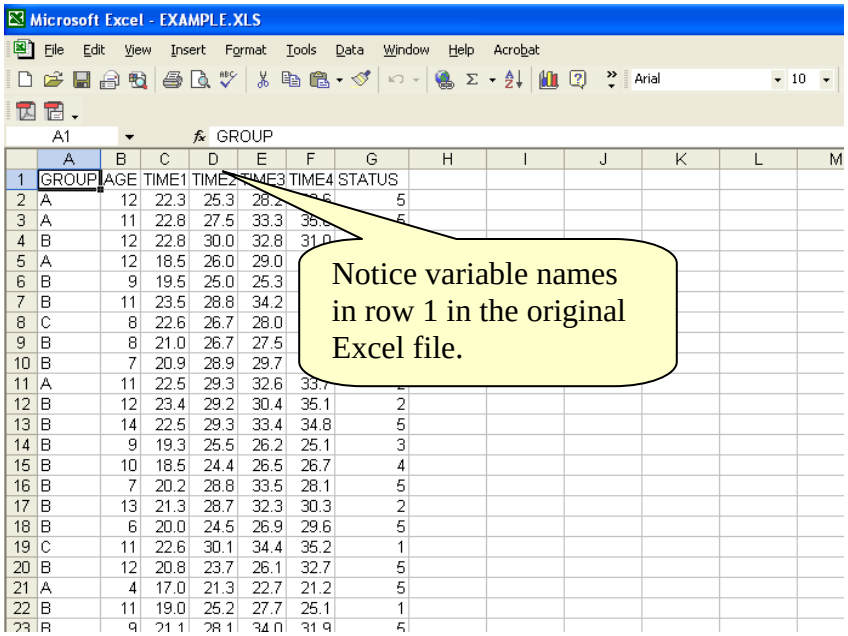
Experiment with other options in Excel, but keep in mind that you can easily distort the message of a chart by applying inappropriate enhancements to the plot. Once the plot is in Excel you can use all Excel options to copy, save, and print the plot.

Important: The Excel program continues to run unless you close it. If you want to save the graph (and data associated with the graph) select **File/Save** as in the Excel Program.

Step 5: End the Excel program by selecting **File/Exit** to return to the main SDA data grid.

Tutorial 5 – Opening an Excel data file

SDA can open files in several common data formats including Microsoft Excel. The data to be imported must be in an acceptable format. For example, an Excel file in your SDA data directory named EXAMPLE.XLS contains the following information:



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	GROUP	AGE	TIME1	TIME2	TIME3	TIME4	STATUS						
2	A	12	22.3	25.3	28.2		5						
3	A	11	22.8	27.5	33.3		5						
4	B	12	22.8	30.0	32.8	31.0							
5	A	12	18.5	26.0	29.0								
6	B	9	19.5	25.0	25.3								
7	B	11	23.5	28.8	34.2								
8	C	8	22.6	26.7	28.0								
9	B	8	21.0	26.7	27.5								
10	B	7	20.9	28.9	29.7								
11	A	11	22.5	29.3	32.6								
12	B	12	23.4	29.2	30.4	35.1	2						
13	B	14	22.5	29.3	33.4	34.8	5						
14	B	9	19.3	25.5	26.2	25.1	3						
15	B	10	18.5	24.4	26.5	26.7	4						
16	B	7	20.2	28.8	33.5	28.1	5						
17	B	13	21.3	28.7	32.3	30.3	2						
18	B	6	20.0	24.5	26.9	29.6	5						
19	C	11	22.6	30.1	34.4	35.2	1						
20	B	12	20.8	23.7	26.1	32.7	5						
21	A	4	17.0	21.3	22.7	21.2	5						
22	B	11	19.0	25.2	27.7	25.1	1						
23	B	9	21.1	28.1	34.0	31.9	5						

Notice that this file includes variable names in the first row and that each column contains consistent data (text or numeric entries). A few guidelines for preparing your data in Excel for use in SDA are:

- The first row of the Excel file should contain SDA compatible variables names (must start with an alpha character and contain no blanks).
- Each column must contain consistent data values. For example, in an AGE column (which should be numeric) you should not include any non-numeric data such as >29 or NA.
- Character codes such as M and F for male and female should be in consistent case.
- Avoid blank cells in your data set. If you have missing values use a missing value code such as -9 for AGE (an impossible value).
- Once you've imported your data set, be sure to define any missing value codes in SDA (Attributes tab)

For a more detailed description of how to prepare your data file in Microsoft Excel, see Elliott et al (2006). (See references.)

The following is an example of how to open an Excel file in SDA. You do not need to have Excel on your computer for this to work.

Importing an Excel File into WINKS

Step 1: Select **File/Open Dataset.**

Step 2: In the Open dialog box select **File of Type** and choose **XLS – Excel 97 or newer.**

Step 3: An example file on disk that you can use for this example is named EXAMPLE.XLS. Select **EXAMPLE.XLS** and Open.

Step 4: You are asked to select the worksheet to import. (This dialog box contains additional information about importing Excel Files.) Select the "Example" worksheet and click Import.

Step 5: You are asked if the first row contains variable names. Choose **Yes.**

Step 5: The file is opened. It is a good idea to save the file in SDA format. Select **File/Save as** and choose Files of type SDA – and Save.

Step 6: If needed, click on the Attributes tab and define any needed Missing values or adjust the names, types and widths of each variable.

The data are now ready to use in an SDA analysis. We recommend that the save the file as an .SDA file.

Reference: Elliott AC, Hynan LS, Reisch JS, Smith JP. "Preparing Data for Analysis Using Microsoft Excel." *Journal of Investigative Medicine.* Vol 54. No. 6, September 2006.

To order WINKS SDA, go to the website (and click on the Purchase button)

www.texasoft.com

To get a copy of this document, and for other WINKS tutorials, go to

http://www.texasoft.com/tutindex.html