

WINKS SDA 7

Statistical Data Analysis and Graphs

"Windows KwikStat"

Getting Started Guide

Version 7

(For BASIC and PROFESSIONAL Editions of WINKS SDA)

www.texasoft.com

© TexaSoft, 2011

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.

Updates

To make sure you have the latest version of WINKS, you may install the latest update at: (Follow the instructions on the screen.)

<http://www.texasoft.com/updates>

First time: When you begin WINKS SDA for the first time, a dialog box appears, asking for an **activation code**. Enter the code provided with your CD. (Some vendors provide the code as a “Serial Number.”) **Enter the code to activate WINKS.** Otherwise WINKS will run in trial mode. Installation places a WINKS icon on your desktop. **Click the WINKS icon on your desktop** (or choose WINKS from your Program menu) to begin the program.

The first time you run WINKS, you are prompted to select program options. We recommend that you **click Ok** to accept the initial program defaults. You can change the options later by choosing "Change Setup Options" from the Help menu. For more information, see *Changing Setup Options* in the *WINKS Reference Guide*.

WINKS
HINT

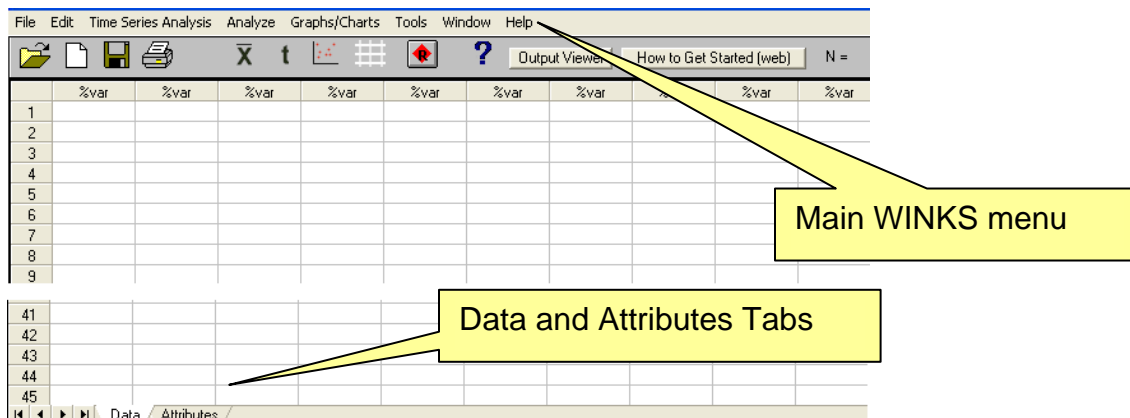
For a quick start using WINKS do these tutorials first.

Tutorial 1 –Using an Existing Data Set

(Learning about WINKS menus)

This first tutorial describes several features of WINKS SDA and shows you how to open an existing data set, calculate descriptive statistics on a single variable, and display a histogram. Follow these steps:

Step 1: Begin the WINKS SDA program. The Main WINKS screen, also called the “data grid” screen appears. A part of the screen is shown here:



Note the two highlighted items – the main WINKS menu and the Data and Attributes tabs.

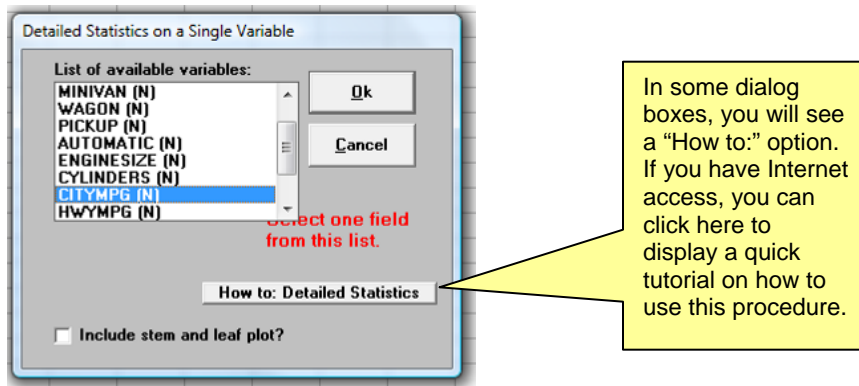
Step 2: From the **File** menu, select “**Open Data Set...**” (a menu selection sequence will be noted in bold (for example **File, Open Data set...**) in future examples. When you select this menu option, a list of data (.SDA) files appears. Open the file named **CARMPG.SDA**. The following data is displayed in the data grid (shown partially below).

The screenshot shows the WINKS SDA main screen with the data grid populated. The columns are: BRAND, MODEL, MINIVAN, WAGON, PICKUP, AUTOMATI, ENGINESIZ, CYLINDERS, CITYMPG. The rows are numbered 1 through 8. A yellow callout box points to the CITYMPG column with the text 'A column of data is called a “Variable” or “Field” and is designated by a name such as CITYMPG.'

	BRAND	MODEL	MINIVAN	WAGON	PICKUP	AUTOMATI	ENGINESIZ	CYLINDERS	CITYMPG
1	TOYOTA	Prius	0	0	0	1	1.50	4	60
2	HONDA	Civic Hybrid	0	0	0	1	1.30	4	48
3	HONDA	Civic Hybrid	0	0	0	1	1.30	4	48
4	HONDA	Civic Hybrid	0	0	0	0	1.30	4	48
5	HONDA	Civic Hybrid	0	0	0	0	1.30	4	48
6	VOLKSWAG	Golf	0	0	0	0	1.90	4	48
7	VOLKSWAG	Jetta	0	0	0	0	1.90	4	48
8	VOLKSWAG	New Beetle	0	0	0	0	1.90	4	48

Step 3: To calculate common statistics on a variable (a column of data) such as *CITYMPG* (City Miles per Gallon) select **Analyze, Descriptives, Detail Statistics/One Variable** from the main WINKS menu. A dialog box appears where you select which variable to use in the calculations.

To select a variable for the analysis, **click on the name (*CITYMPG*)** in the list of variables, and click **Ok**.



Step 4: The output from the calculations is displayed in the Output Viewer and contains descriptive statistics for the *CITYMPG* variables (abbreviated here.)

Variable Name is CITYMPG

N	= 1,081	Missing or Deleted = 0
Mean	= 19.28955	St. Dev (n-1) = 4.60479
Median	= 19.00	St. Dev (n) = 4.60265
Minimum	= 10.00	S.E.M. = 0.14005
Maximum	= 60.00	Variance = 21.20405
Sum	= 20852.00	Coef. Var. = 0.23872
Skewness	= 2.157	Kurtosis = 10.579

Continues...

Extreme values (Case Number)

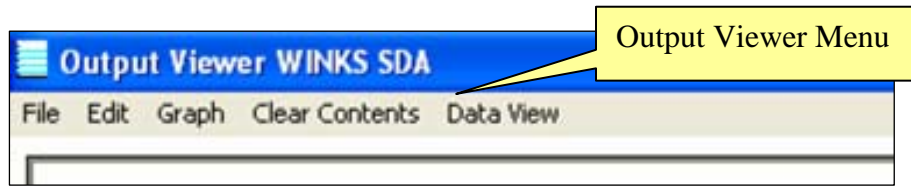
LOWEST 5	HIGHEST 5

10.00(1081)	60.00(1)
11.00(1079)	48.00(2)
11.00(1080)	47.00(3)
12.00(1077)	46.00(4)
12.00(1076)	45.00(5)

This output includes a number of statistics for *CITYMPG* including the mean, standard deviation, N (sample size), confidence intervals (not shown here) and the 5 lowest and 5 largest values (helpful to locate extreme values (outliers).)

Notice for this example that the lowest value for *CITYMPG* was 10, for car number 1081 in the data set. The highest *CITYMPG*, 60, is for car number 1.

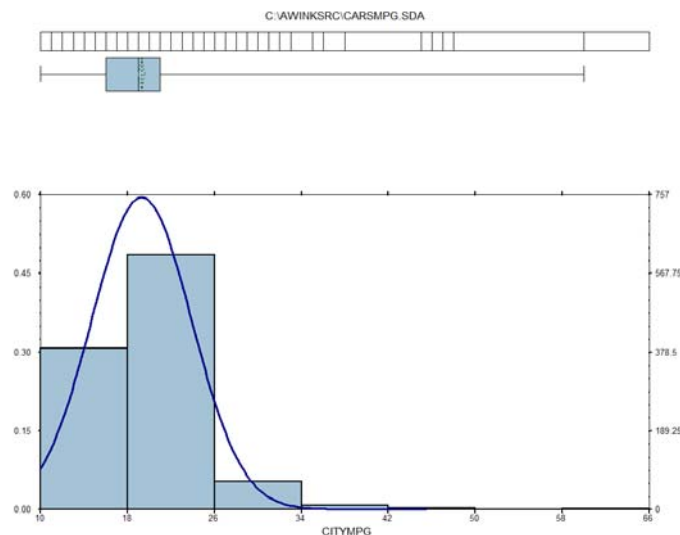
Step 5: The WINKS Output Viewer menu is shown here.



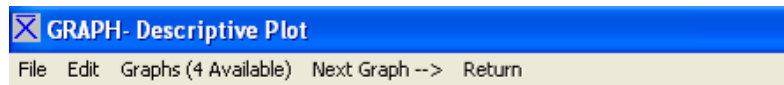
The menu items for the Output viewer are:

- **File** – Save output to a file as RTF (Rich Text Format), TXT (text) or DOC (Microsoft Word), Open the file in MS Word (requires that you have MS Word installed), or print.
- **Edit** – Select All, Copy, and/or Paste.
- **Graph** – Display a graph appropriate to the analysis.
- **Clear Contents** – Clear the contents of the Output Viewer. (Handy if you are using the Append option from the program setup.)
- **Data View** – Return to the main WINKS screen (data grid.)

Step 6: Display a graph of this data by selecting (from the Output Viewer menu) **Graph, Display Graph**. A screen appears showing a histogram of the *CITYMPG* data.



Right-click on the plot to display an options dialog box. In this dialog box, check the "Display normal curve" option. You can also change other aspects of the plot such as the title, fonts, axes, etc. When you finish changing items in the Options dialog box, **click OK** to redisplay the graph (as shown above.) The menu items for this graph window are shown here and described:



File - Print or save the graph.

Edit - Select plot options, or to capture the plot to the clipboard.

Graphs (4 Available) – Select from a list of graphs.

Next Graph - For this analysis, 4 graphs are available. Click “**Next Graph →**” to view the next graph. The four graphs are:

- Histogram with box plot and optional bell curve
- Cumulative distribution
- Cumulative normal
- Normal Probability Plot

Return: Return to the WINKS Output Viewer.

Step 7: Return to the Output Viewer by clicking **Return** or by selecting **File, Exit**.

Step 8: In the Output Viewer, click **Data View** to return to the Main WINKS window, the data grid screen.

WINKS
HINT

Doing the tutorials first is time well spent – in about 30 minutes you will learn most of what you need to know to use WINKS SDA.

Tutorial 2 – Using WINKS SDA GRAPHS

(Learning about WINKS graphs)

This tutorial shows how to display graphs without first performing an analysis. Graphs are of two types:

- **WINKS graphs** – Graphs designed to help you understand the results of statistical tests within WINKS.
- **Excel graphs** – Graphs displayed in Microsoft Excel (you must have Excel installed for this option to work.)

See later chapters for more graph examples.

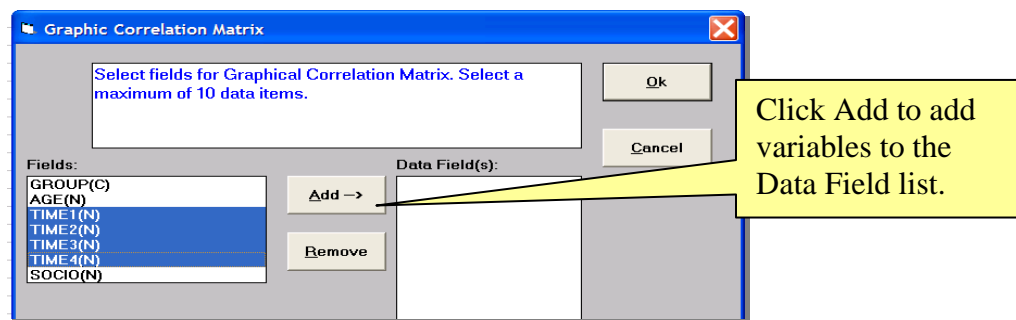
WINKS Graph Example 1 (Scatterplot matrix)

Display the relationship between several pairs of quantitative (continuous) variables in a single graph.

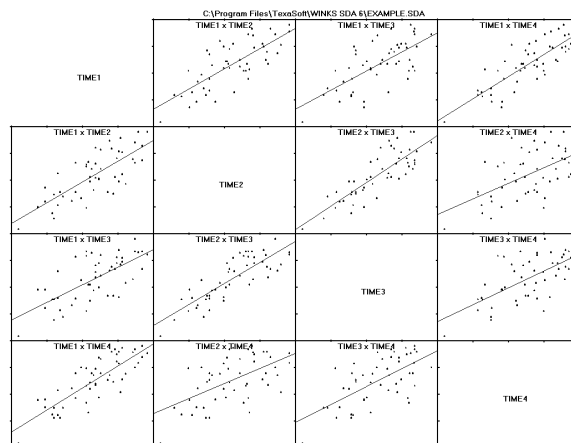
Step 1: On the main WINKS data screen, select **File, Open data set...** and open the **EXAMPLE.SDA** data file.

Step 2: Select **Graphs/Charts, Correlation Matrix of X-Y Scatterplots.**

Step 3: In variable selection dialog box, select *TIME1* to *TIME4* by highlighting the variables names (on the left) and click **Add**.



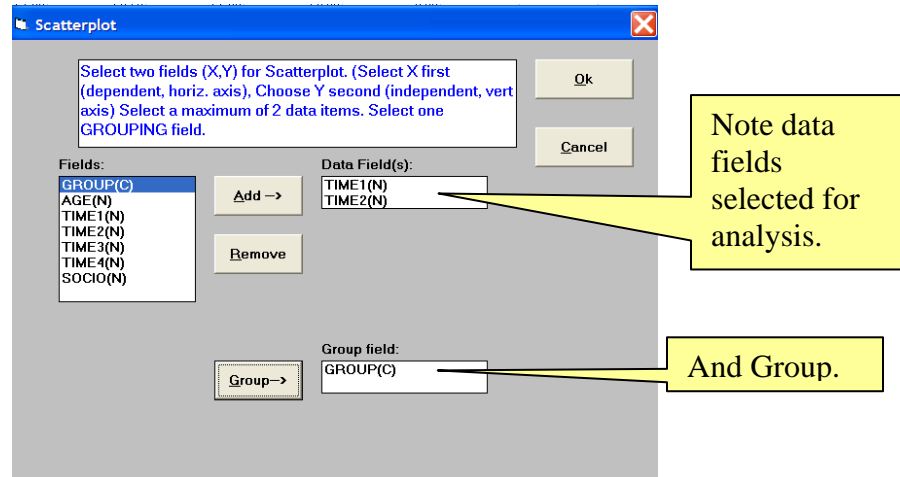
Step 4: Click **Ok** and a matrix of graphs is displayed. Select **Edit, Display Regression Lines** to display the regression lines shown on the graph. This plot shows the relationship between two variables at a time. A cloud of points close to the line indicates a stronger relationship than a highly dispersed cloud of points.



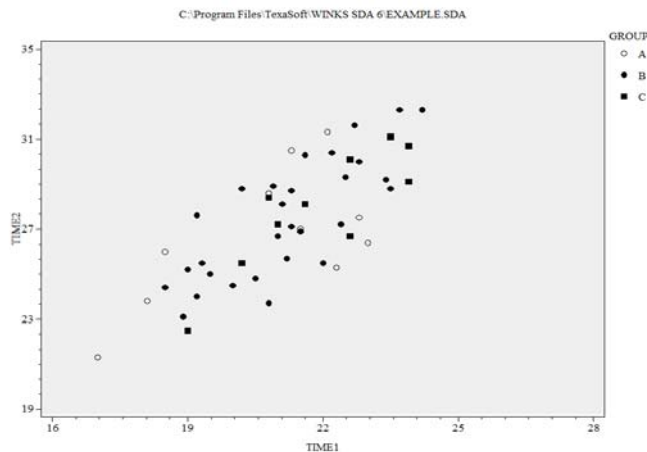
Step 5: Return to the WINKS main screen by clicking the **Return** menu option or by selecting **File, Exit**.

WINKS Graph Example 2 (Single scatterplot with groups):

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

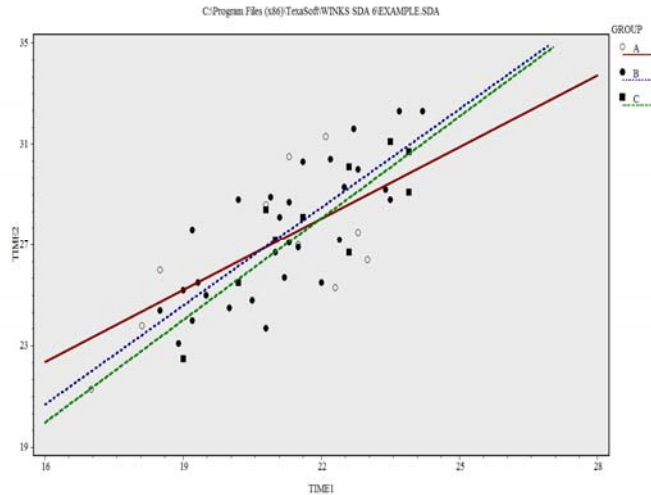


Click **Ok** to display the following graph:



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

Step 2: Select **Edit, Options** (or right-click on the edge of the window) to display the options dialog box. **Click the Graph tab** and **check Display Regression Line(s)** and click **Close**. 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** (or **Return**) to return to the data screen.

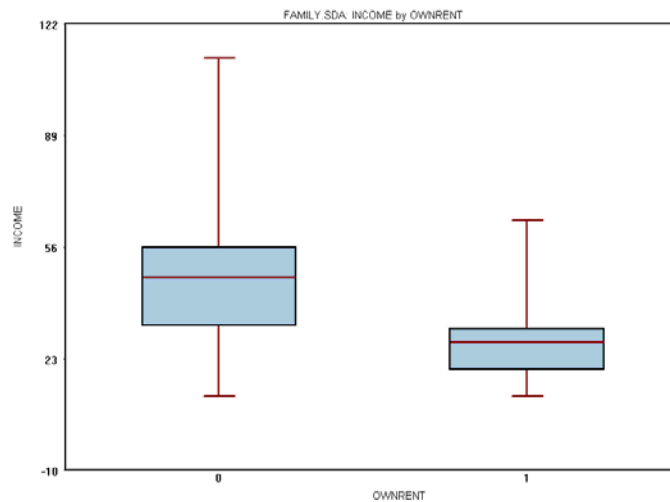
WINKS
HINT

You can also change colors, fonts, etc. for this plot using the options menu.

WINKS Graph Example 3: (Group Comparison Plots)

Step 1: Select **File, Open data set...** and open the data set named **FAMILY.SDA**. Select **Graph/Charts, By Group Plots (Box Plots, etc.)**

Step 2: Select *Income* as the Data field and *OwnRent* as the Group Field. Click **Ok** to display the comparative box plots:

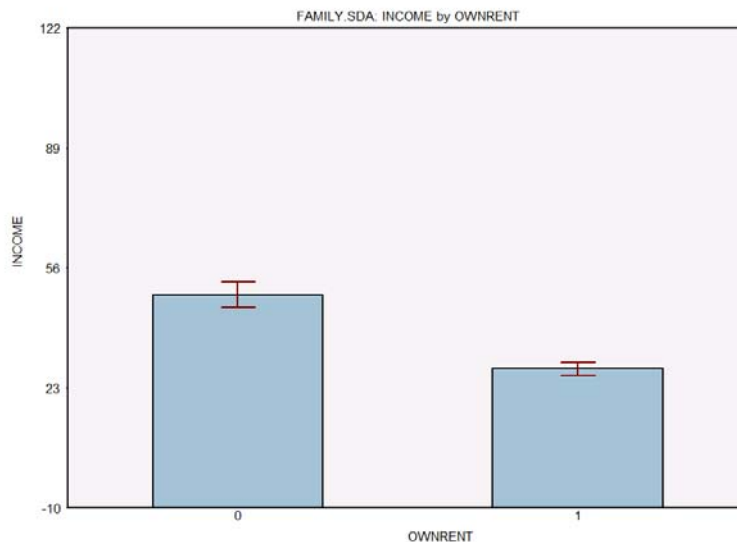


The box plots show the distribution of data. The horizontal lines represent the median for each group (own or rent house) and the box contains 50% of the data. It is not surprising that the income for group 0 (owns house) is larger than for those who rent.

Step 3: Right click anywhere on the plot to display plot options.

- Deselect “Box and Whiskers Plot”
- Select “Error Bars”
- Select “Bar Graph of Means”

Optionally change box color by clicking “Select box color” **Click Ok** to redisplay the plot using the selected options.



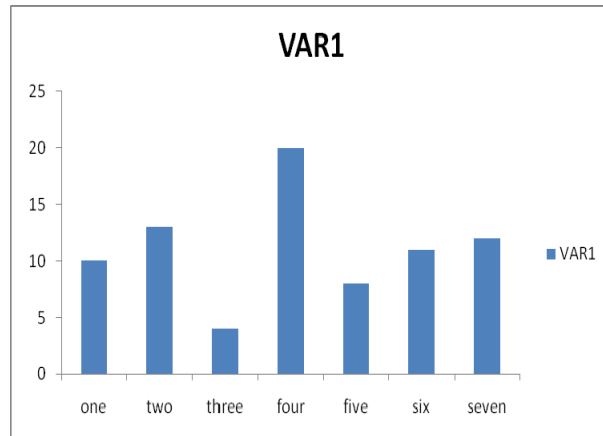
Click **Return** to return to the data sheet.

Excel Graph Example 1: Illustrates how to create a bar chart in Excel from summarized count data in WINKS. (Requires that Excel be loaded on your computer.) Some Excel options vary according to which version of Excel you are using.

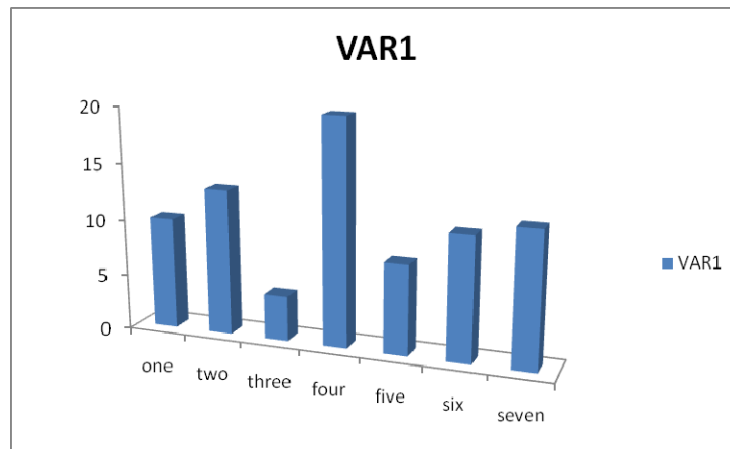
Step 1: Open **BARChart.SDA**.

Step 2: Select **Graphs, Bar Charts, Excel Bar Chart (from summarized data)**.

Step 3: Click on *VARI* in the field box and **Add**. Click on *LABEL* variable name in the Field box and click on the **Label** button. Click **Ok**. Excel opens a worksheet and bar chart as shown below. (If the graph does not appear, look at the program bar at the bottom of your screen. You may have to click on the Excel icon to display the Excel graph.)



Step 4: Right click on Graph and from the displayed Excel menu change the chart type to a **3-D vertical (column) chart** type (menus differ in different versions of Excel) to display the graph in a 3D format as shown below:



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.

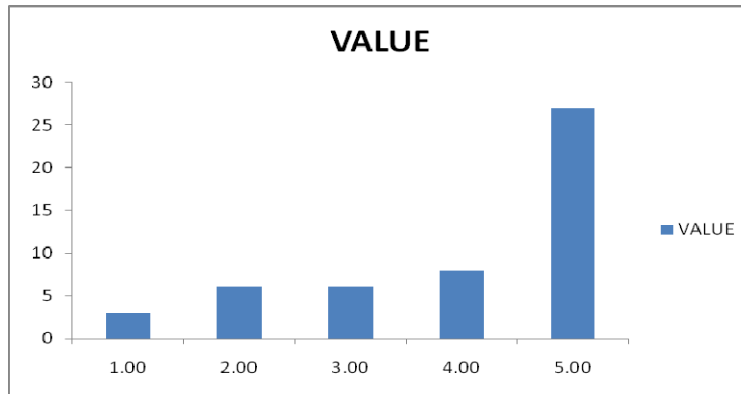
Step 5: End the Excel program by selecting **File, Exit**. You are prompted to click a "Continue" button to return to WINKS.

Excel Graph Example 2: Illustrates creating a bar chart in Excel from raw counts.

Step 1: Open **EXAMPLE.SDA**.

Step 2: Select **Graphs, Bar Charts, Excel Bar Chart (from frequencies)**.

Step 3: Select **SOCIO_STATUS (Socioeconomic Status)** as the data value and click **Ok** to display the Excel chart shown here.



Step 4: Close Excel to return to WINKS.

See **Part 4, Graphs and Charts** in the *Reference Guide* for more graph capabilities.

Tutorial 3 – Analyzing Summary Data

(Illustrates entering summary data for analysis)

The following examples illustrate how to perform an analysis from summary data when data are in the form of means or counts.

Summary Data Example 1 (Comparing two means): Illustrates how to perform a two-sample Student's t-test (comparing two means) using summary data.

A quality of life (QOL) score is measured on patients undergoing cancer treatment. Subjects are randomly assigned to one of two support groups. Group 1 uses written instructions on what to discuss and Group 2 is lead by a facilitator. After six weeks, a survey is used to measure QOL. High scores indicate more satisfaction (scale is 0 to 30): The results are:

Data for Group 1 (Written instructions)

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

Data for Group 2 (Facilitator)

Mean = 27.97, Standard Deviation = 3.22, N = 9

Compare these means using an Independent Group Student's t-test:

Step 1: From the **Analyze** menu, select **t-tests and ANOVA, Ind. Group from Summary Data**. (No data set has to be opened.)

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

Name	Mean	St. Dev.	Sample Size
Written	23.44	3.41	8
Facilitator	27.97	3.22	9

Step 3: Click **Ok** to display the results. (abbreviated here.)

Group Means and Standard Deviations

```
-----
Written: mean = 23.44      s.d. = 3.41      n = 8
Facilitator: mean = 27.97  s.d. = 3.22      n = 9
```

Mean Difference = -4.53 S.E. Difference = 1.61417

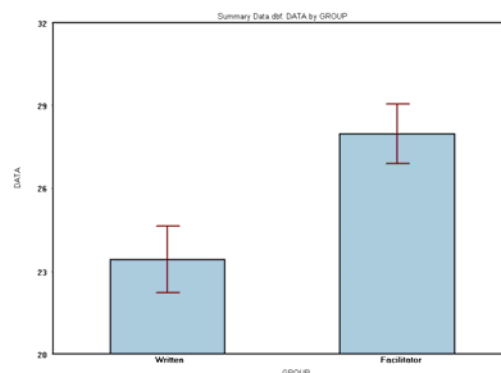
Independent Group t-test on Summary Data

```
-----
Equal variance: Calculated t= -2.82 with 15 D.F. p = 0.013 (two-tail)
```

The test of interest is in the table labeled “Independent Group t-test on Summary Data.” The p-value ($p = 0.013$) indicates sufficient evidence (based on the small probability) that the satisfaction scores for the two groups are different. This provides evidence that the QOL mean score for the Facilitator group ($M = 27.97$) is statistically larger (better) than for the Written group ($M = 23.44$).

Note: For a more thorough interpretation of the output, see “t-Tests and ANOVA” in Part 4 of the *Reference Guide*.

Step 4: To display a plot of the comparison, select **Graph, Display Graph**. The following plot provides a visual confirmation that the mean QOL score for the Facilitator group is larger than for the “Written” group.



You can select options for this plot as described in the previous tutorial to edit this plot.

Step 5: Return to the Output Viewer by clicking **Return** on the graph menu. From the Output Viewer, return to the main WINKS screen by clicking on **Data View**.

Summary Data Example 2 (Comparing proportions in a 2 x 2 table): Illustrates how to compare proportions across groups using count data. The follow table is from historical data on smoking and cancer.

	Disease Status	
	Had Lung Cancer	No Lung Cancer
Smoking Habit		
No Smoking to Slight Smoking	56	956
Moderate to Excessive Smoking	269	1646

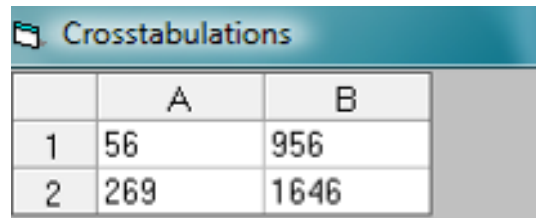
To test the hypothesis that there is no relationship between smoking and cancer, follow these steps:

Step 1: From the **Analyze** menu, select **Analyze from Summary Data, Chi-Square Table/Summary Data**. (No data set has to be opened.)

Step 2: You are prompted to select output options. **Check** the following options - Frequencies, Total Percent, Row Percent and Total Percent. Click **Ok**.

Step 3: Enter the number of rows and columns in the data table. In this case **Rows=2** and **Columns=2**. Click **Ok**.

Step 4: A blank table is displayed with the appropriate number of rows and columns. Enter the counts for the table as shown here:



	A	B
1	56	956
2	269	1646

Step 5: Optionally, enter labels in the text boxes at the bottom of the data entry screen as shown below. (The entry “Smoking, No, Yes” tells the program to label the row as “Smoking” with the categories “No” and “Yes”).

Row label:	Smoking, No, Yes
Col label:	Cancer, Yes, No

Step 6: Click **Calculate** to display preliminary test results at the bottom of the screen. Click **Exit** to go to the Output Viewer. The following output (abbreviated) is displayed. The count data and percents are shown in a crosstabulation table. Notice the total, row, and column percents in each cell.

Smoking(rows) by Cancer (columns)

FREQUENCY			
TOT PCT			
ROW PCT			
COL PCT	Yes	No	TOTAL

No	56	956	1012
	1.9	32.7	34.6
	5.5	94.5	
	17.2	36.7	

Yes	269	1646	1915
	9.2	56.2	65.4
	14.0	86.0	
	82.8	63.3	

TOTAL	325	2602	2927
	11.1	88.9	100.0

The number 5.5 indicates the Row Percent of 56 subjects out of 1012 total. (56/1012=0.55)

The test to determine if there is an association between Smoking and Cancer in these data is the Chi-Square test, as reported in the following table of the output:

Statistic	DF	Value	p-value
Chi-Square	1	48.617	<0.001
Yates' Chi-Square	1	47.758	<0.001
Fisher's Exact Test (one-tail)			0.000
(two-tail)			0.000

Chi-Square p-value.

Because the probability of no association is small ($p < 0.001$), you reject that hypothesis and conclude that there is a statistically significant association between smoking and lung cancer in this data. **Click Graph, Display Graph to view a chart for these data. Click Return and Data View** to return to the WINKS data grid. For a more complete interpretation of this test, see Part 4 of the *Reference Guide*.

Other Analyses from Summary Data in WINKS SDA

Other analyses in WINKS SDA that utilize summary data include:

- Independent Group Means Comparison (t-tests and Analysis of Variance)
- Single Sample t-test
- Dunnett's Test (ANOVA with pairwise comparisons to a standard.)
- 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
- Bar and Pie Charts
- Mantel-Haenszel (Meta Analysis) (Professional Edition)
- Kappa – Interrater reliability (Professional Edition)
- Multiple comparisons from table (Professional Edition)
- Pie and bar charts
- P-value determination
- Probability calculations (Combinations and Permutations)
- Forest Plots /Meta Analysis (Professional Edition)

Tutorial 4 – Editing Data in WINKS

(Creating a data set)

This tutorial provides several examples about how to manipulate data in the WINKS data editor. For more information, see Part 3 of the *Reference Guide*.

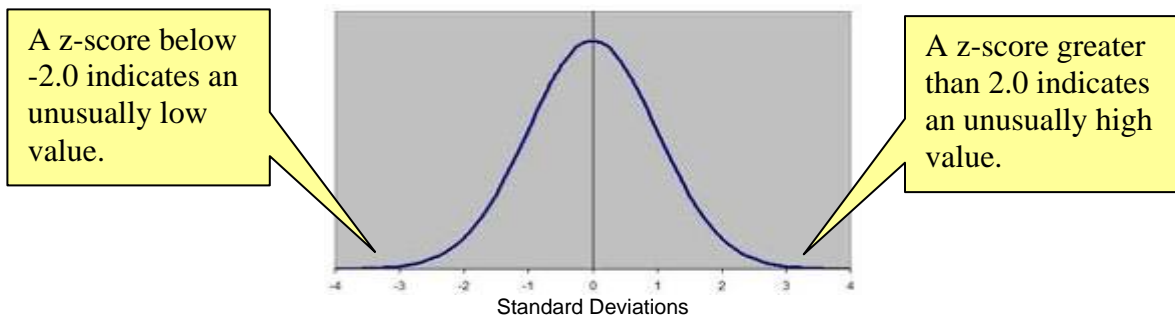
Data manipulation Example 1: Create a new variable (z-score). This example illustrates how you can create new variables from the information in a current data set.

Step 1: Open the data set CITY.SDA. This data set contains demographic information about 480 American cities.

Step 2: Transform the INCOME data into a z-score: A z-score transforms the data into a standard normal variate with a mean of 0 and a standard deviation of 1 using the formula

$$z = \frac{(x - \bar{x})}{s}$$

where x is an observation in your sample data, \bar{x} is the mean and s is the standard deviation. The distribution of the resulting z-score is illustrated in this standard normal (bell-shaped) curve:



When data is transformed to a standard z-score, you can quickly see what observations represent unusually high or low scores by noting values greater than 2.0 (standard deviations from the mean) or less than -2.0 (standard deviations from the mean.)

On the WINKS data grid screen, select **Edit, New Variable, Common Transformations**. In the dialog box combo list, select **z-score**. Specify **-99** as the **missing value code** (A missing values code is an impossible number -- see an explanation of missing values in Part 3 of the *Reference Guide*.) **Click Ok**.

Step 3: Select INCOME from the list of variables and click **Ok**.

	NAME	INCOME	TAX	UNEMPLOYMN	POP80	POP70	WEATHER	HOUSEPRICE	Z_INCOME
1	Akron, OH	23682	791	1	660328	679239	2	55750	-0.54965
2	Atlantic City, NJ	23525	1728	1	194119	175045	1	52228	-0.59432
3	Baton Rouge, LA	25310	282	0	493973	375628	3	47114	-0.08646
4	Bridgeport, CT	33518	1708	0	395455	401752	1	58107	2.24885
5	Buffalo, NY	23749	2507	1	1242573	1349211	2	64292	-0.53058
6	Chattanooga, TN	23479	616	1	426540	370857	2	48927	-0.60740
7	Denver, CO	30545	1054	0	1619921	1239545	3	78112	1.40299
8	El Paso, TX	18411	885	1	479899	359291	1	48147	-2.04933
9	Gary, IN	26668	828	1	642781	633367	3	48434	0.29992
10	Green Bay, WI	23614	929	0	175280	158244	3	51897	-0.56899

Step 4: Notice that a new column is added to the data set, named *Z_INCOME*. This column contains the z-transformed data. In this data, notice that the z-score for Bridgeport, CT is 2.24885. This indicates that this city has a comparatively high income and the score for El Paso, -2.04933, indicates that it has a comparatively low income.

Once you calculate a new variable using the “**New Variables by Common Transformations**” option, you can use those variables in analyses and save the data set to retain the new information. See Using the Edit menu in the *Reference Guide*, Part 2 for additional information about transformations.

Data manipulation Example 2 (recode): Create a new variable by Recode: Recode *HOUSEPRICE* in the *CITY* dataset into three groups: HIGH, MODERATE and LOW.

Step 1: Select **Edit, New Variable by Recode**.

The screenshot shows the 'Recode' dialog box with the following configuration:

- Enter name of new variable for recoded item:** COST
- Recode values from this field (Select field):** HOUSEPRICE
- Recode result as numeric, or as character:** as character
- For HOUSEPRICE values between:**

"GT or EQ to"	and	"Less Than"	Recode as...
0		50000	LOW
50000		90000	MODERATE
- ELSE recode value as:** current value or code as HIGH
- Note:** Comparisons are >= left number and < right number specified.
- Recode MISSING in field as:** MISSING

Step 2: In the Recode dialog box, enter the name of the new variable as *COST*. Select *HOUSEPRICE* as the variable to use for the recode, and select “as character” for the recode.

Step 3: Indicate how the recode will be performed.

- Enter the range 0 to 50000 and the recode value “LOW” in the first recode row

- Enter the range **50000 to 90000** and the recode value **“MODERATE”** in the second row. (Do not use commas in these numbers.)
- In the ELSE category, enter **“HIGH.”**
- Enter **“MISSING”** in the missing textbox (although there are no missing values in this data set.)

Step 4: Click Ok. Observe that a new character-type variable is created – with values LOW, MODERATE or HIGH corresponding to the values of *HOUSEPRICE*. (You can also use recode for numeric recodes.) The first few records look like this:

Note new recoded variable “COST.”

	NAME	INCOME	TAX	UNEMPLOYMN	POP80	POP70	WEATHER	HOUSEPRICE	COST
1	Akron, OH	23682	791	1	660328	679239	2	55750	MODERATE
2	Atlantic City, NJ	23525	1728	1	194119	175045	1	52228	MODERATE
3	Baton Rouge, LA	25310	282	0	493973	375628	3	47114	LOW
4	Bridgeport, CT	33518	1708	0	395455	401752	1	58107	MODERATE
5	Buffalo, NY	23749	2507	1	1242573	1349211	2	64292	MODERATE
6	Chattanooga, TN	23479	616	1	426540	370857	2	48927	LOW
7	Denver, CO	30545	1054	0	1619821	1239545	3	78112	MODERATE
8	El Paso, TX	18411	885	1	479899	359291	1	48147	LOW
9	Gary, IN	26668	828	1	642781	633367	3	48434	LOW
10	Green Bay, WI	23614	929	0	175280	158244	3	51897	MODERATE
11	Houston, TX	32565	723	0	2905350	1999316	3	65738	MODERATE
12	Jacksonville, FL	23428	521	0	737519	621827	2	50605	MODERATE
13	Kalamazoo, MI	25434	1298	1	279192	257723	2	48278	LOW

NOTE: There are a number of other edit menu options you can use to manipulate your data set to get it in the form you need for analysis. For more information, see Part 3 of the *Reference Guide*.

Step 5: Once you edit records using the Edit menu options, **save the data** set to retain your changes. These variables are now available to use for any analysis.

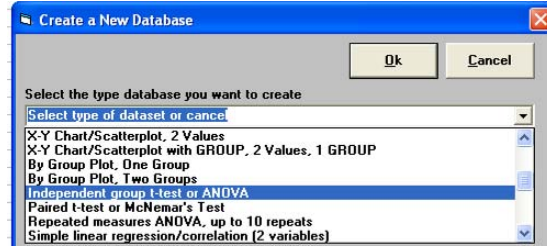
Note: Calculations and recodes used to create new variables in a WINKS data set are a one-time event. If you change values in the dataset, these new variables are not automatically recalculated.

Tutorial 5 – Creating a data set and entering data

(Compare means of two groups)

Creating a Data Set, Example 1: A plant nursery owner is interested in which of two fertilizers causes trees (from seed) to grow the tallest (in cm) in six weeks. For this experiment, 13 plants are randomly assigned to 2 fertilizers and grown under identical conditions. For this example, code the fertilizers as 1 or 2 (This is an arbitrary grouping name -- you could also use names such as ACME and DELUX.) Enter the data using these steps:

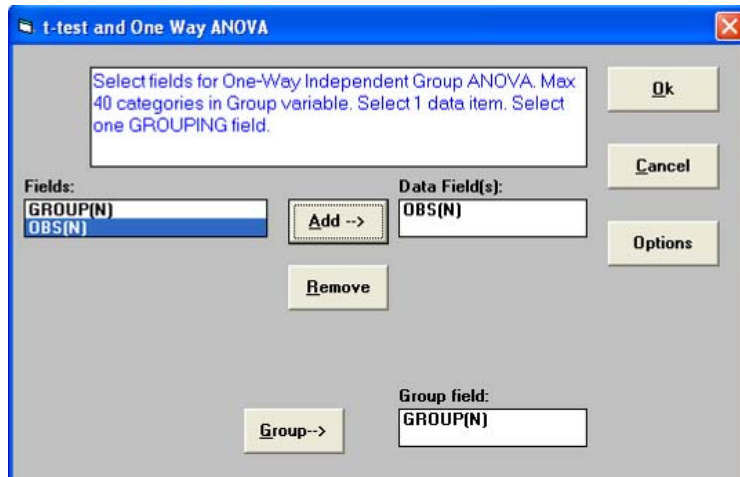
Step 1: Create a data set by selecting **File, New Data set....** A list of analysis types appears as shown:



Since each plant is a unique “subject” and the comparison of two means is a t-test, **select the option “Independent group t-test or ANOVA”** and click **Ok**. A blank data sheet is displayed with columns named *GROUP* and *OBS*. **Enter the following data** in the appropriate columns of the data grid.

<u>GROUP</u>	<u>OBS (HEIGHT)</u>
1	46.2
1	55.6
1	53.3
1	44.8
1	55.4
1	56.0
1	48.9
2	51.3
2	52.4
2	54.6
2	52.2
2	64.3
2	55.0

Step 2: Once data are entered, save the data by choosing **File/Save As...** and enter TUTORIAL5 as the data set name. To perform the analysis, select **Analyze, t-test and ANOVA, Independent Group (t-test/ANOVA)**. The following dialog box appears:



Click OBS in the Fields list on the left, and **Add→** to place *OBS* in the Data Fields box. **Click GROUP** in the left box and click **GROUP →** to place *GROUP* in the Group field box.

Step 3: Click **Ok** to display the results (abbreviated here):

Group Means and Standard Deviations

```
-----
1: mean = 51.4571          s.d. = 4.7476          n = 7
2: mean = 54.9667          s.d. = 4.7944          n = 6
```

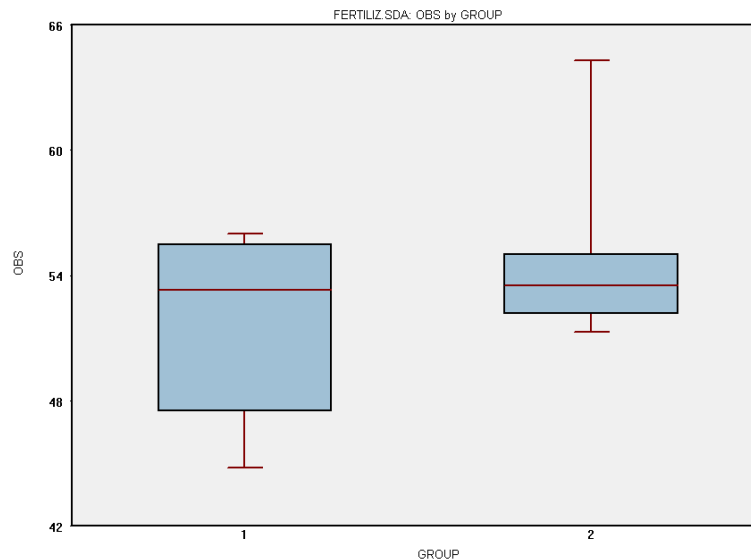
```
Mean Difference = -3.50952      S.E. Difference = 2.65538
```

Independent Group t-test on OBS

```
-----
Equal variance: Calculated t= -1.32 with 11 D.F. p = 0.213 (two-tail)
```

For this data, the p-value for the Student's t-test is $p=0.213$. Since the p-value is greater than 0.05, you do not have enough evidence to reject the null hypothesis that the two means are equal. Therefore, you conclude that there is not enough evidence to claim that one fertilizer is better than the other. For more information on interpreting the t-test, see Part 4 of the *Reference Guide*.

Step 4: Click on the Graph menu in the Output Viewer to display a group comparison graph as shown here:

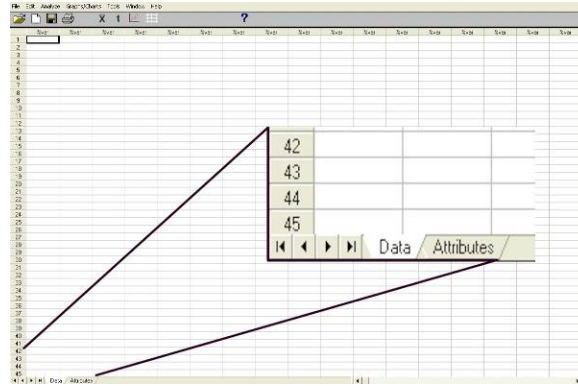


The default graph is a box and whiskers plot. You can change the type of plot, colors, and other options by right clicking at the edge of the window or selecting Edit/Options.

Creating a Data Set, Example 2: If the data set type you want to create is not in the pre-set list, you can design your own data set using the following steps. This example creates a similar data set to the one used in Example 1 (the fertilizer example.)

Step 1: Select Data New... and Cancel when prompted to select a type of data set. On the main WINKS SDA screen, notice the tabs in the lower left corner (Data and

Attributes.) (To create a new data set, you can immediately enter data into the data grid, but we recommend that you follow this procedure.)



Step 2: Click the Attributes tab at the bottom of the WINKS SDA grid to display an Attributes screen:

	Name	Type	Width	Decimals	Missing	Label	Formats
1	SUBJECTNO	C (Character)	10	0			
2	BEFORE	N (Number)	10	1			
3	AFTER	N (Number)	10	1			
4							
5							

Step 3: Define your data set variables. Suppose you have data on 8 subjects that represent weight before and after a six-month diet.

- In the Name column enter *SUBJECTNO* in row 1, *BEFORE* in row 2 and *AFTER* in row 3.
- In the *SUBJECT* row, select (C) Character as the data type. For *BEFORE* and *AFTER*, select (N) Number.
- The WIDTH column indicates how many maximum characters to allow for the data, and the Decimals column allows for decimals for numeric variables. Enter the information as shown above.
- Leave the Missing, Label, and Format columns blank for now.

Step 4: Click the Data tab to return to the data grid. Notice that the first three columns are names "*SUBJECTNO*" and "*BEFORE*" and "*AFTER*." Now you can enter the data shown here:

SUBJECT	BEFORE	AFTER
0001	162	168
0002	170	136
0003	184	147
0004	164	159
0005	172	143

0006	176	161
0007	159	143
0008	170	145

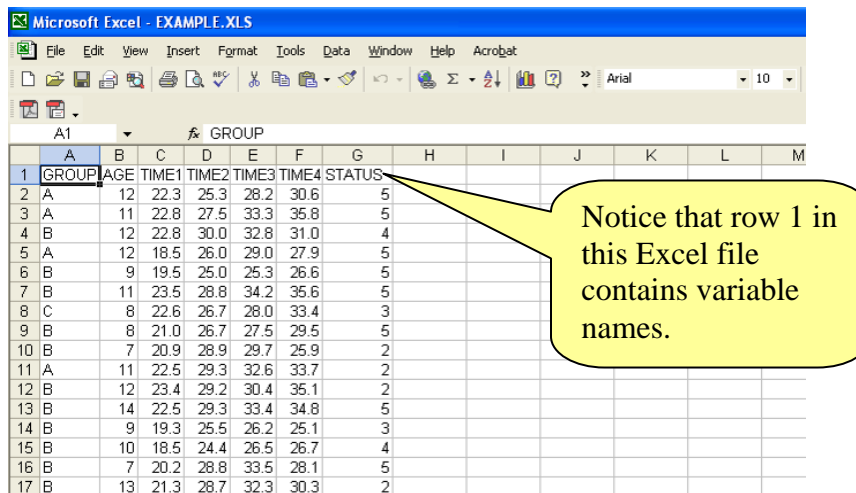
Step 5: (Optional). Analyze this data as a paired t-test by **selecting Analyze, t-test and ANOVA, Paired/Rep. Measure**. Select *BEFORE* and *AFTER* as the data fields and click Ok to display the results. The t-test ($t(7) = 3.70687$, $p = 0.0076$) indicates a significant average weight loss of 19.375 pounds.

Note: You can edit the information in the Attributes at any time – for example if you didn't leave enough room for decimal places, you can change that, or return and enter missing value codes, etc.

Tutorial 6 – Opening an Excel data file

(Use similar procedures for .txt, csv and dbf file types.)

WINKS SDA can open files in several common data formats including Microsoft Excel (.xls), Text (.txt), Comma Separated Values (.csv), and dBASE (.dbf). 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	30.6	5						
3	A	11	22.8	27.5	33.3	35.8	5						
4	B	12	22.8	30.0	32.8	31.0	4						
5	A	12	18.5	26.0	29.0	27.9	5						
6	B	9	19.5	25.0	25.3	26.6	5						
7	B	11	23.5	28.8	34.2	35.6	5						
8	C	8	22.6	26.7	28.0	33.4	3						
9	B	8	21.0	26.7	27.5	29.5	5						
10	B	7	20.9	28.9	29.7	25.9	2						
11	A	11	22.5	29.3	32.6	33.7	2						
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						

Excel file with data prepared for import into WINKS

For guidelines on preparing an Excel Data file so you can open it in WINKS, see Part 3 of the *Reference Guide*.

Using Excel Data Example 1: Open an Excel File in WINKS:

Step 1: Select **File, Open data set...** Depending on your version of Windows, the Open dialog box allows you to select “Files of type” to open. **Select the .xls Excel file type.** Files with an .xls filename extension are displayed in the list of files.

Step 2: Select the file named **EXAMPLE (EXAMPLE.XLS)** and click **Open**.

Step 3: WINKS prompts you to select which worksheet in the Excel file to import. For this example, **accept the default and click Import**.

Step 4: WINKS asks if the first row of the worksheet contains variables names. **Select Yes**.

The data are imported into the WINKS data grid. You can now use the data for any analysis. We recommend that you

1. Save the data set as a WINKS .SDA file
2. Look at the Attributes tab to verify that the data have been imported correctly (as far as numeric and character variables)
3. Add missing value codes if needed.

Note: Your Excel worksheet must be in .XLS format (not .xlsx format) for WINKS to open it. If it is in the .xlsx format, resave it in Excel using "save as type" into the .xls format. Or, use method 2 described below.

Using Excel Data Example 2: Copy and Paste from Excel. You can copy and paste information from any version of Excel file into WINKS. (or from any other program that has data in a table format, including web pages.) If you have any problems opening an Excel file in WINKS, use this technique:

Step 1: Open EXAMPLE.XLS in Excel (any version.) **Highlight and copy the data cells** (not the column names). **Go to WINKS**, and with your **cursor at the top left** of a blank data sheet, select **Edit, Paste**. The data are copied into the WINKS data grid.

Step 2: Click on the Attributes tab to name the variables and specify data types, and missing values. Data are now ready for use. As in the previous example, we recommend that you

1. Save the data set as a WINKS .SDA file
2. Look at the Attributes tab to verify that the data have been imported correctly (as far as numeric and character variables)
3. Add missing value codes if needed.

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

Tutorial 7 – Advanced Topic: Labels and Formats Tutorial

This tutorial illustrates how to define variable labels and format labels.

VARIABLE LABEL -- A "Variable Label" is an alternate (and usually more explanatory) description of a variable. For example, a label for AGE might be "AGE in 1990," and for ID, "Identification Number." A Label makes the output more readable.

VARIABLE FORMAT - A "Variable Format" is a series of labels describing categories. For example, for the variable ANSWER with values 0 and 1. Define 0="No" and 1="Yes." When output is generated, No and Yes are used rather than the cryptic 0 and 1.

Step 1: Open **QUESTIONS.SDA**. Click on the **Attributes** tab. The following screen is displayed. Note the last 2 columns -- Label and Formats.

	Name	Type	Width	Decimals	Missing	Label	Formats
1	ID	N (Number)	10	0			
2	AGE	N (Number)	10	0			
3	GENDER	N (Number)	10	0			
4	Q1	N (Number)	10	0			
5	Q2	N (Number)	10	0			
6	Q3	N (Number)	10	0			

Label and
Formats

Step 2: To define a label for *ID*, click on the **Label** cell on the same row as *ID*, and enter **Identification**. For *AGE* enter **AGE in 2011**. You could also enter labels for the *GENDER* and *Q1* through *Q20* if desired. The *Q#* variables are the answers to 20 Yes-No questions on a survey.

Step 3: To define a format for *Q1*, click on the **Formats** column, *Q1* row in the Attributes grid. The following screen appears:

In the text box next to label "IF Q1 =" enter the number 1. Next to the label "then assign label =" enter **No**. Click the button "Add category label to list," and the format description

1 = No

appears in the large text box. Similarly, enter the number 2 next to "IF Q1 =" and **Yes** next to "then assign label =" and click the button "Add category label to list," and the format

2 = Yes

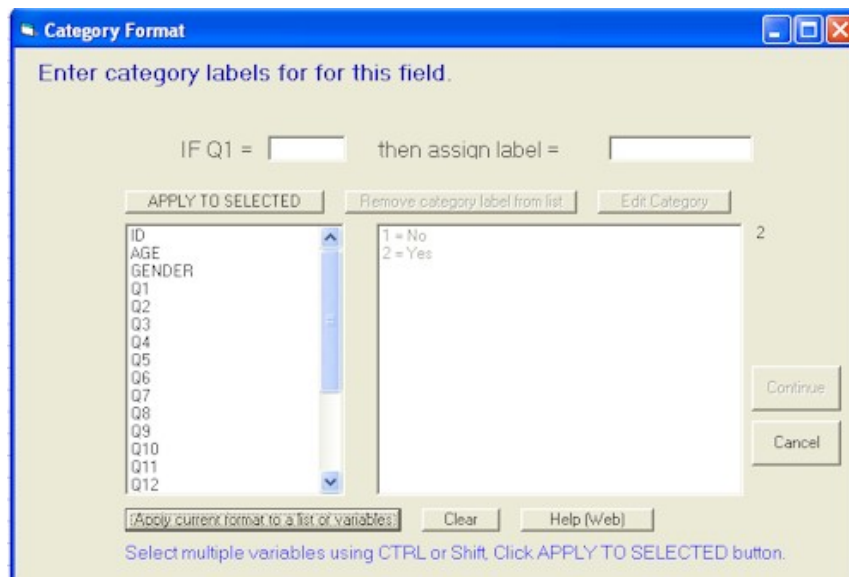
appears in the large text box. Click the **Continue** button to assign this format to the *Q1* variable.

Step 4: Assign the same format to a number of variables. Click on the format for *Q1*, and the dialog box appears, with the defined format

1 = No

2 = Yes

in the large text box. To assign this same format to the other questions, click on the button "Apply current format to a list of variables." The dialog box changes to the following:



Step 5: Select variables from the left text box list. In this case select *Q2* through *Q20*. You can use the Ctrl or Shift to select multiple variables at a time. For example, click on *Q2*, hold down Shift, then click on *Q20* to select *Q2* through *Q20*.

Step 6: Click on the "APPLY TO SELECTED" button. You are notified that the format has been applied to the select variables.

Step 7: Click Continue. The attributes screen appears. Verify that the format has been copied to the selected variable rows.

Step 8: Click the Data tab at the bottom of the screen. Note that the format does NOT change any values in the data sheet. However, when you perform an analysis (Such as a frequency) the format names will appear. For example, select the menu option

Analyze/ Crosstabs, Frequencies, Chi-Square/ Frequencies

and specify the variables *Q1* through *Q20*. The following (abbreviated) output is displayed:

Frequency Table for Q1

Q1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NO	40	80.0	40	80.0
YES	10	20.0	50	100.0

Frequency Table for Q2

Q2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NO	48	96.0	48	96.0
YES	2	4.0	50	100.0

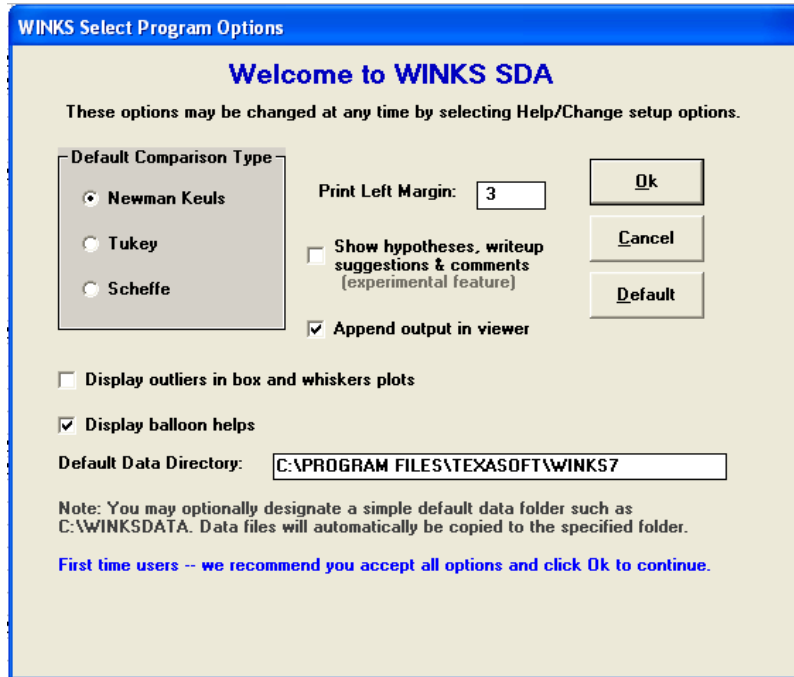
Frequency Table for Q3

Q3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NO	45	90.0	45	90.0
YES	5	10.0	50	100.0

Notice that the format for the answers for the "Q" variables are displayed as Yes and No rather than 1 and 2.

Changing Program Options

During the initial program setup, the “Select Program Options” dialog is shown. You can also display and change these options at any time by choosing “Change Setup Options” from the Help menu.



Here are brief explanations of the program option items:

- **Default comparison type:** For analysis of variance, this selects the default multiple comparison procedure used. (You can also select comparison type when selecting variables for an analysis. See Part 4, One-Way ANOVA.)
- **Show hypotheses:** causes additional information to be displayed for many analyses including the hypothesis tested, write-up suggestions, and comments.
- **Append output:** causes output from each analysis to be appended to the end of the previous analysis. Otherwise, the results of an analysis are cleared in the output viewer before the results of the current analysis is displayed.
- **Display outliers:** causes default box and whiskers plots to include indications for outliers.
- **Display balloon helps:** Causes yellow “help” boxes to display when checked (or not when not checked) for icon options on main menu page
- **Default data directory:** If you do not want your data stored in the default data directory, enter a new directory path here and examples files will be copied to the new location.

Which Edition, BASIC or PROFESSIONAL?

- **Reference Guide.** This WINKS SDA Reference Guide includes information about features in *both* the BASIC and PROFESSIONAL Editions. Parts 1, 2 and 3 are relevant to both editions.
- **WINKS SDA BASIC Edition:** provides commonly used data analysis procedures for researchers, analysts and students – statistical procedures usually covered in a one-semester college course. See the Reference Guide Contents, **Part 4 (BASIC Edition Procedures)** for a list of procedures included in the BASIC Edition.
- **WINKS SDA PROFESSIONAL Edition:** includes all of the BASIC features *plus advanced statistical procedures*. See the Reference Guide Contents, **Part 5 (Professional Edition Features)** for a list of procedures included in the PROFESSIONAL Edition.
- **Updating from BASIC to PROFESSIONAL.** To upgrade your WINKS SDA from the BASIC to PROFESSIONAL edition, go to the following website:

<http://www.texasoft.com/upgrade/>

or contact TexaSoft sales at info@texasoft.com

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

www.texasoft.com

To get a copy of this document, other WINKS tutorials, and a link to the full *WINKS Reference Guide*, click on the Tutorials button from the main TexaSoft webpage.