User Manual

XL Data Analyst™

Professional Edition

Version 3.0

November 2024

Table of Contents

Windows Versus Mac

In the vast majority of cases, XL Data Analyst Pro will work on Mac-based Excel without any problems.

XL Data Analyst Pro was developed and has existed as a Microsoft Windows Macro-enabled program for a great many years. Only very recently has Microsoft or Apple made it possible to run macro-enabled Microsoft Excel programming on Mac's. Consequently, while XL Data Analyst Pro has had extensive development and testing on Windows, conversion to and testing on Mac is far less extensive.

On the plus side, after considerable revision, practically everything XL Data Analyst Pro on Mac looks very similar to the Windows version. The analysis findings worksheets are, in fact, identical. Further, no special provisions, configurations, or settings are necessary to run XL Data Analyst Pro on a Mac.

With the provided datasets, all analyses run smoothly and with identical results on a Mac. However, occasional crashes occur on a Mac for the XL Data Analyst Pro "Clean-Up" and "Filter" operations. While XL Data Analyst has a built-in Fatal Error Catch (see: "Troubleshooting"), Mac has its own fatal error recovery system that is allowed with these operations. In all testing, the Mac recovery completely resolves the fatal error when the user is prompted to recover the work. So, except for a slight disruption in flow, no work is lost.

If users collectively experience significant issues with XL Data Analysis Pro on Mac or on Windows, efforts may be made to resolve them.

XL Data Analyst Pro Quick Start

Hey, we know that you don't want to read a big manual to use XL Data Analyst Pro. So here is a quick start resource for you.

Download It

Download any XL Data Analyst Pro file from the Download page on its website. Save the file on your hard drive, pin drive, or some other storage location.

Try It

Use Excel to open up the XLDA file. Enable macro if prompted. The datasets for XL Data Analyst Pro are for tryout. Use the XL Data Analyst Pro menu now on Excel to do whatever analyses you want to check it out. Save your work either with its original name or one of your choosing.

Adapt It

When you are sufficiently oriented and feel that it will be useful, import your dataset as a comma-separated-variable or Excel worksheet file into XL Data Analyst Pro. Then, on the Define Variables worksheet, set up the Description, Value Codes, and Value Labels for each variable. Use Clean-Up and make necessary corrections. Do whatever XL Data Analyst Pro analyses you want and save your work as a Macro-enabled (.xlsm) Excel file with a unique file name. The next time you start up XL Data Analyst Pro, open up your file and resume work with it.

Introduction

XL Data Analyst Pro is a data analysis system developed for use with Microsoft Excel (runs on PC-based Excel 2016/Version 16 and later versions). It performs data analyses of various types and generates tables and graphs with professional appearance that can be copied into other applications such as Microsoft Word or PowerPoint. As compared to other statistical analysis programs and systems, XL Data Analyst Pro provides "interpreted" output. That is, instead of requiring the user to inspect computed statistical values such as F, t, or z, it provides a "plain English" statement of the analysis finding based on the user's desired level of confidence. Thus, XL Data Analyst Pro is more user-friendly than, for instance, SPSS, and it minimizes interpretation errors due to unfamiliarity, faulty recall, or other difficulties with traditional statistical analysis output.

A significant development is that XL Data Analyst Pro Version 3.0 runs on both Windows and Mac pc's and laptops. In the past, it did not run on Mac's. While this manual describes the Windows use of XL Data Analyst Pro, the Mac use looks very similar.

The "Pro" (Professional) edition of the XL Data Analyst incorporates a number of features that are useful to contemporary researchers. These improvements include:

- Excel ribbon for the XL Data Analyst Pro
- Optional elimination of outliers
- Specification of effect sizes
- Data visualization of significant findings
- Use any significance level specified by the user
- Output arranged by effect size and significance
- Discovery analysis (exploratory data analysis)
- Progress bar
- Format options for tables and graphs
- Data filter feature
- Export workbook feature
- Variables display

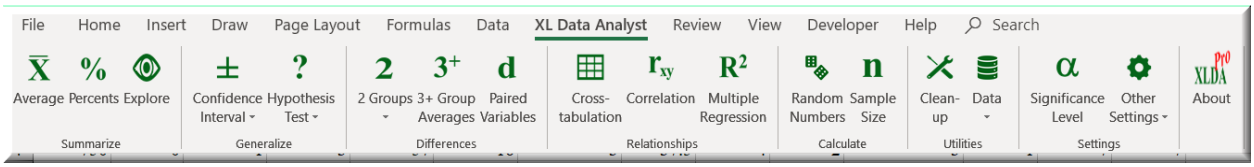
Quick Background. Originally, the XL Data Analyst (Basic) was developed as a part of ***Basic Marketing Research***, editions 1-3, a marketing research textbook authored by Alvin Burns and Ronald Bush (Prentice Hall, 2012, 3rd edition). XL Data Analyst Pro as a stand-alone analysis tool kit that does not require deep knowledge of statistical procedures. Currently, XL Data Analyst Pro Version 3.0 is used in ***Marketing Research***, 10th edition, by Al Burns and Ann Veeck (Pearson, 2025). Innumerable improvements and tweaks have taken place as a result and especially in ensuring that it will run on Mac pc's and laptops.

XL Data Analyst Pro is designed for maximum user-friendliness and ease of use. Whereas anyone may use it, it is primarily a pedagogical tool for instructors of marketing research. For analyses that utilize statistical tests, XL Data Analyst Pro displays ***interpreted*** output, meaning that the desired level of confidence is applied, and the findings are explained in plain English. Thus, XL Data Analyst Pro is useful for a range of commonly-used univariate analysis procedures such as descriptive analysis, generalizations, differences tests, and crosstabulation. Apart from basic multiple regression, the XL Data Analyst does not perform multivariate analyses such as factor analysis, discriminant analysis, cluster analysis, or other such analyses.

Basic Operation of XL Data Analyst Pro

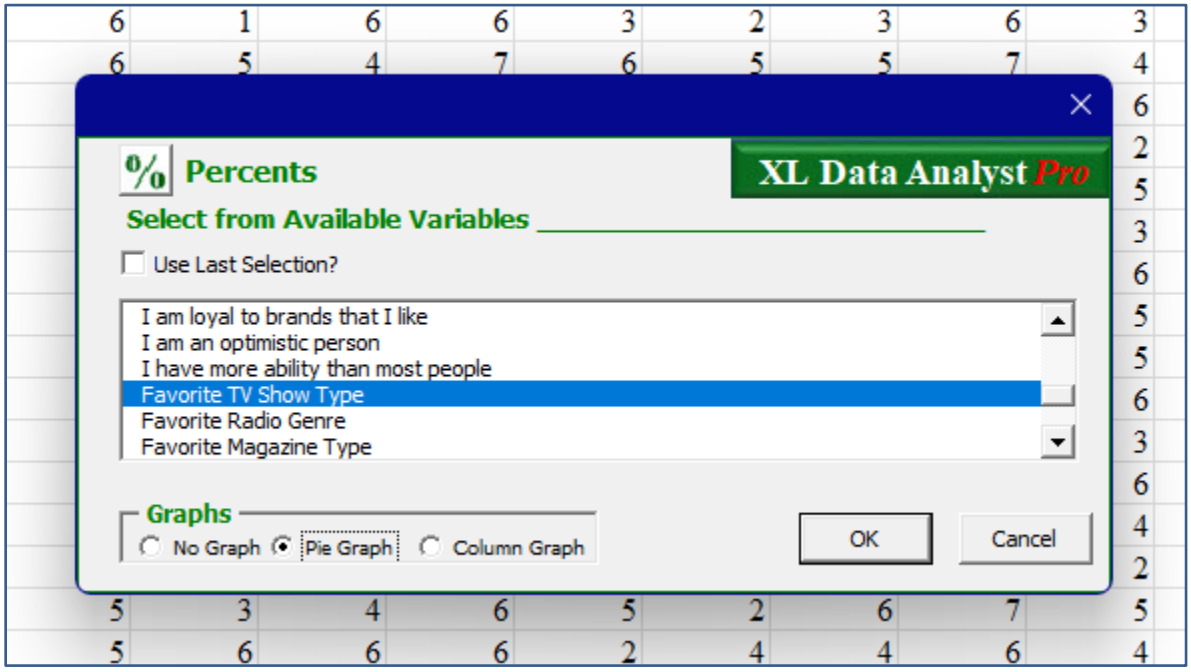
All analyses in the *XL Data Analyst Pro* utilize a simple three-step procedure.

1. Select the analysis from the *XL Data Analyst Pro* Excel ribbon menu. A click on any analysis will direct XL Data Analyst Pro to open its corresponding variable(s) selection window.



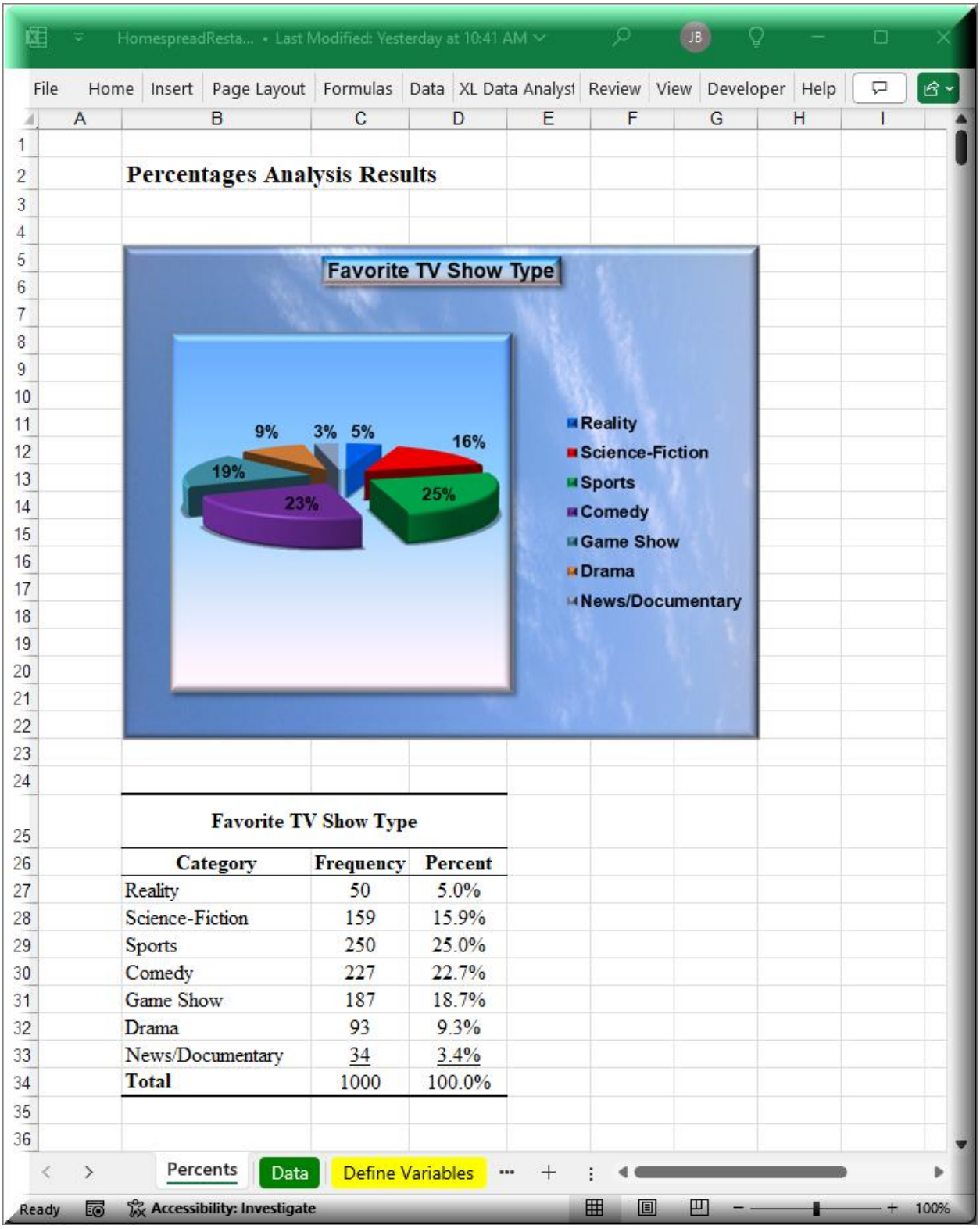
XL Data Analyst Pro Ribbon Menu

2. Select the variables to analyze in the *XL Data Analyst Pro* selection window that appears. In some windows, users may be required to enter values (e.g. performing a hypothesis test requires that the hypothesis number be entered). Most analyses have options such as removal of outliers, arrangement of output, ability to recall last used variables, and so on.



Example Analysis Variable Selection Window

3. Examine and interpret the output that appears on the resulting output worksheet. If significance testing is called for, the output is arranged with significant (confidence level)/largest (effect size) findings prominent and visual presentations of these findings.



Example XLDA Pro Output

As can be seen in the figure above, an analysis result is placed on a new worksheet that is provided with a descriptive name. Subsequent analyses of the same type are placed on new worksheets that are numbered sequentially, such as Percents1, Percents2, and so on. Users can rename analysis results worksheets, if desired.

Start-Up and Running XL Data Analyst Pro

XL Data Analyst Pro is a macro-enabled Excel file (.xlsm) that runs on a Windows or a Mac pc or laptop with Microsoft Excel Version 16 or higher. It will **not** run on Microsoft Excel Online nor on tablet-based Excel because these platforms disallow Excel macros.

There is no installation other than copying a file onto a storage device of the user's choice. This procedure can be accomplished by downloading the file from the XL Data Analyst Pro website using the following steps.

Start up and running XL Data Analyst Pro is simple.

1. Go to the "Download" page of the XL Data Analyst Pro website and click on the download button for the XL Data Analyst Pro file.

2. With the download function, use "Save as..." and identify the location at which you wish to save the XLDA Pro .xlsm file.

3. Start-up and use XL Data Analyst Pro is as follows:

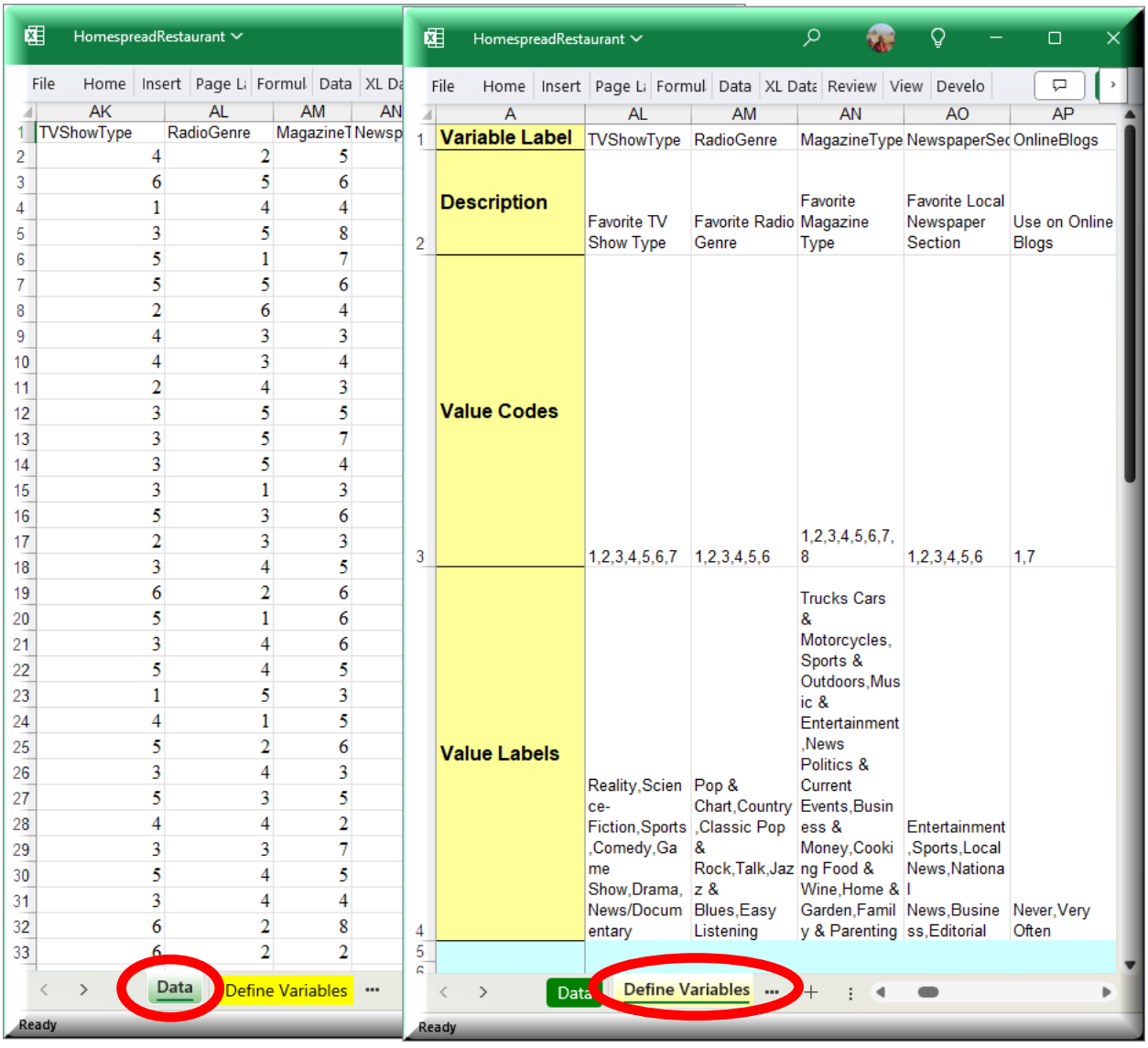
    a. Open Excel.
    b. In Excel, use the File-Open option.
    c. Find the XL Data Analyst Pro .xlsm file in your saved location and click on it. (Or: just click on the XL Data Analyst file to open it with Excel)
    d. Enable/Allow macros if prompted by Excel.
    e. Perform analyses with the XL Data Analyst Pro. All analyses will generate additional worksheets in the XL Data Analyst workbook.
    e. Use File-Save to save your work with the original file name. Alternatively, use the "Save as…" function to save it with a new file name of your choosing.

How XL Data Analyst Pro is Organized

This section describes how data are organized, variable descriptions, value codes, and value labels. It also describes how these are related in XL Data Analyst Pro. This information is especially critical for users who utilize their own datasets.

XL Data Analyst Pro has two essential worksheets named "Data" and "Define Variables." The **Data Worksheet** holds raw numbers that constitute the dataset. The Data worksheet is arranged in columns and rows. The columns are associated with variables (such as questions on a questionnaire), while the rows are associated with cases such as respondents or subjects who have answered these questions in a survey. Of course, data sets do not necessarily need to be survey data: any data organized by variables (columns) and cases (rows) may be used. In general, data sets should have only numeric data, without formats such as currency, dates, etc. Data with decimals is perfectly acceptable.

Row 1 of the Data worksheet **must** contain variable labels; although, the "labels" feature of Excel does not need to be invoked. For more detail on the format of the labels on the Data worksheet, refer to "**Using the XL Data Analyst with Your Own Dataset.**"



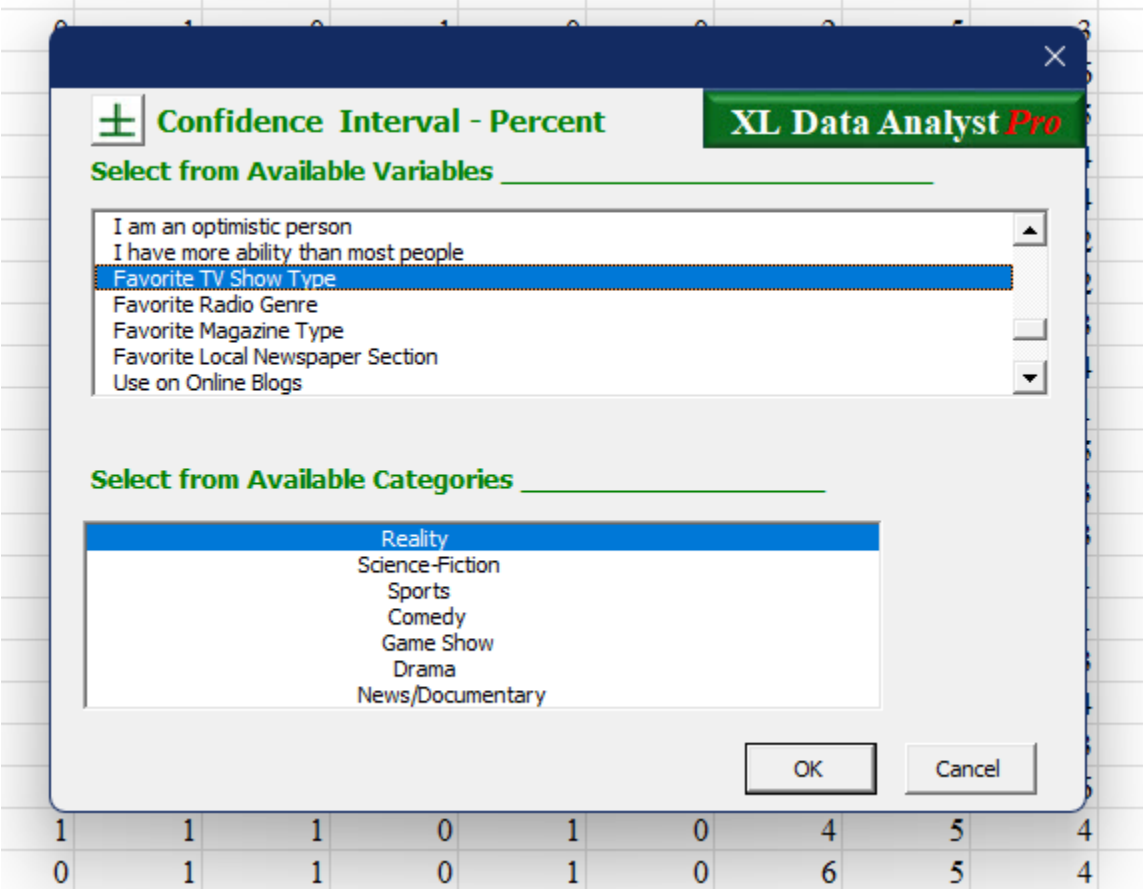
XLDA Data & Define Variables Worksheets

The **Define Variables Worksheet** is set up in parallel with the Data worksheet. As can be seen above, the Define Variables worksheet has the Data worksheet variable labels linked in its Row 1 (lagged by one column). This lag is necessary because protected cells A1-A4 have the terms "Variable Label," "Description," "Value Codes," and "Value Labels" to assist users in understanding what to place in cells B1-B4, C1-C4, D1-D4, etc. if they are building their own XLDA datasets.

Beneath each Variable Label on the Define Variables worksheet is a ***Description*** of that variable of any length desired. The user places the Descriptions in their respective locations on the Define Variables worksheet. Descriptions are vital as they appear in the various XL Data Analysis Pro selection windows, and they appear as identifiers on the output. Also, beneath each Description is a set of ***Value Codes***, or code numbers in the Data worksheet (dataset) that correspond to (e.g.) the answers to the associated question on the survey, while beneath each Value Code cell are the associated ***Value Labels***. Note that Value Codes and Value Labels are separated by commas within their respective cells. The codes and labels pertain to whatever coding system the user has established for the dataset variables.

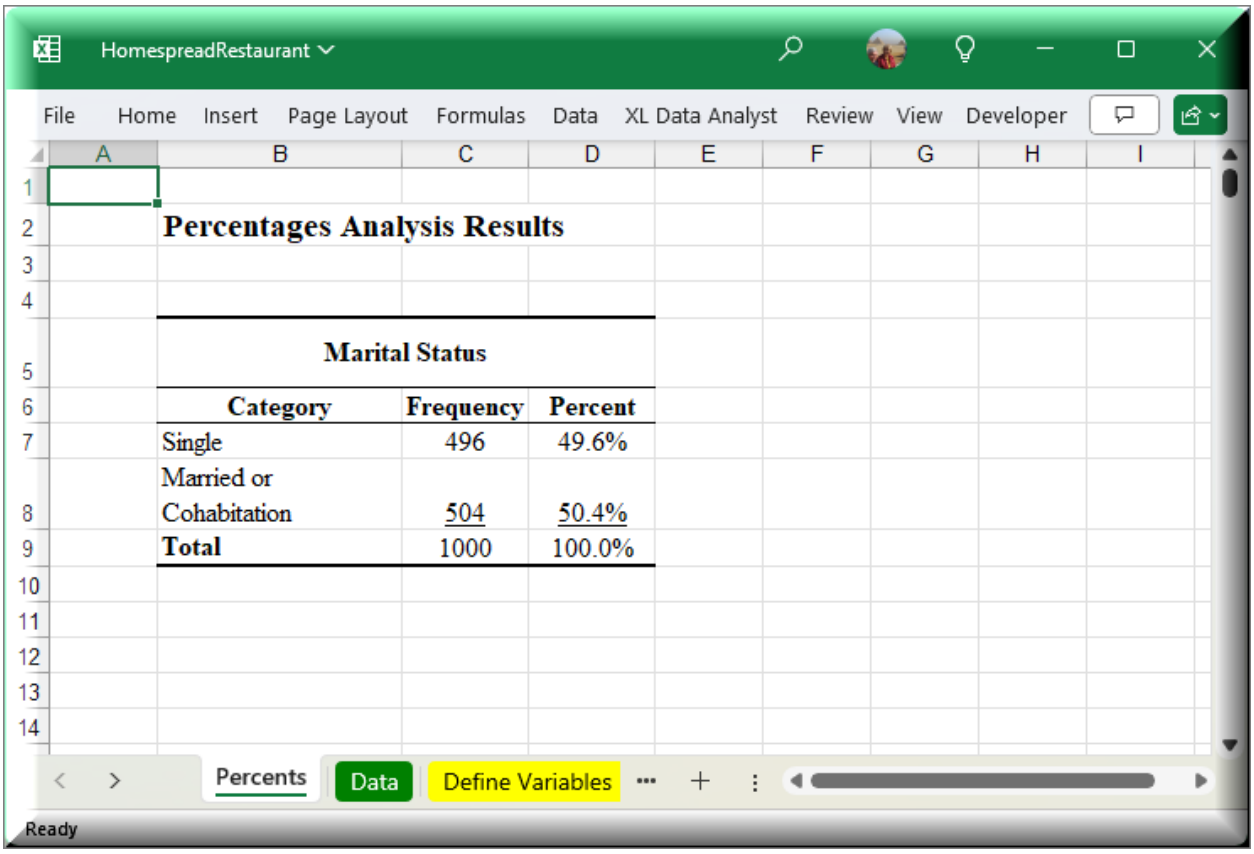
Thus, in the figure above, the Variable "TVShowType" is the label in Row 1 of Column AK on the Data worksheet. It is linked to Row 1 of Column AL on the Define Variables worksheet, and the associated Description is "Size of home town or city." The answer codes are 1,2,3,4,5,6, and 7, and these pertain to the labels, Reality, Science-Fiction, Sports, Comedy, Game Show, Drama, and News/Documentary, respectively. The answer codes and answer labels correspond to the arrangements used in questionnaire design systems such as Qualtrics or SurveyMonkey. However, such systems cannot be imported into XL Data Analyst Pro, so users must input them manually in a one-time set-up operation if the Define Variables worksheet is not set up.

The Value Labels are similarly essential in that they appear in various XL Data Analyst Pro selection windows, and they appear in XL Data Analyst Pro output tables. This screenshot below shows the XL Data Analyst Pro variables selection window for Percentage Confidence Interval analysis. In it the top selection window uses Variable Descriptions (e.g. Favorite TV Show Type, Favorite Radio Genre, etc.) while the bottom window displays the Value Labels (Reality, Science-Fiction, etc.) for the selected variable.



Example XLDA Procedure Window

This screenshot is a Percentage analysis result for Marital Status. The Variable Description is Marital Status, and the Value Labels are Unmarried and Married or Cohabitation.



Example XLDA Output

For detailed information on Variable Labels, Descriptions, Value Codes, Value Labels, and other related topics, please refer to "**Using the XL Data Analyst with Your Own Dataset**"

Datasets Available to Users for Training

Most users will naturally desire to use XL Data Analyst Pro on their own datasets. Consequently, in addition to the dataset contained in XL Data Analyst Pro, there are other datasets users can download and import into XL Data Analyst Pro for the purposes of familiarization and training. This section describes datasets available on the XL Data Analyst Pro website.

These datasets are a mix of developed for ***Basic Marketing Research***, editions 1-3, Pearson, by Al Burns & Ron Bush, those developed for ***Marketing Research***, editions 9 and 10, Pearson, by Al Burns & Ann Veeck, and some potentially interesting free datasets which are somewhat dated but still useful for training.

Users can treat these datasets as training vehicles, inspect them to understand nuances of XL Data Analyst Pro, analyze them for personal enlightenment, or use them as templates for the development of their own XL Data Analyst Pro datasets. With each, information about the dataset is contained on a *Description* worksheet within its XL Data Analyst Pro file.

A description of each dataset follows. Please note that the ".xlsm" extension has been omitted for all datasets.

Homespread Restaurant: This dataset is the one for the integrated case in ***Marketing Research***, 10th edition. The code book for this dataset is evident on the Define Variables worksheet, but the variables, value codes, and value labels are listed here as well.

Variable Description	**Value Codes**	**Value Labels**
Ordered Appetizer on last visit? Ordered Salad on last visit? Ordered Side Order on last visit? Ordered Alcohol on last visit? Ordered Entree on last visit? Ordered Dessert on last visit? Used Applebee's in past month? Used Olive Garden in past month? Used Chili's in past month?	0,1	No,Yes
Price Specials Desirability Online Reservations Desirability Heart-Healthy Dishes Desirability Mixed Drink Specials Desirability Frequent Diner Program Desirability	1-7	1=Very Undesirable,7=Very Desirable
Variety of Choices rating Beverages Selection rating Prices rating Taste rating Presentation rating Portion Sizes rating Preparation rating Accuracy of Order rating Quality rating Waiting Time rating		

Friendliness rating Service Speed rating Attentiveness of Server rating		
Overall Satisfaction with Homespread Restaurant	1-7	Very Dissatisfied,Very Satisfied
Intention to use Homespread Restaurant in the future	1-7	1=Not Likely,7=Very Likely
Frequency of using Homespread Restaurant	1-7	Never,Very Often
I have an active lifestyle I have a healthy lifestyle I like to experience things I am loyal to brands that I like I am an optimistic person I have more ability than most people	1-7	1=Strongly Disagree,7=Strongly Agree
Favorite TV Show Type	1,2,3,4,5,6,7	Reality,Science-Fiction,Sports,Comedy,Game Show,Drama,News/Documentary
Favorite Radio Genre	1,2,3,4,5,6	Pop & Chart,Country,Classic Pop & Rock,Talk,Jazz & Blues,Easy Listening
Favorite Magazine Type	1,2,3,4,5,6,7,8	Trucks Cars & Motorcycles,Sports & Outdoors,Music & Entertainment,News Politics & Current Events,Business & Money,Cooking Food & Wine,Home & Garden,Family & Parenting
Favorite Local Newspaper Section	1,2,3,4,5,6	Entertainment,Sports,Local News,National News,Business,Editorial
Use on Online Blogs Use of Content Communities Use of Social Networks Use of Online Games	1-7	1=Never,7=Very Often
Age		Actual age in years
Highest Education Attained	10,12,13,14,16,18,20	Less than High School,High School Diploma,Some College,Associate Degree,Bachelor's Degree,Master's Degree,Doctorate
Household Income Level	12500,37500,62500,87500,112500,137500,162500	Below $25000,$25000 to $34999,$35000 to $49999,$50000 to $74999,$75000 to $99999,$100000 to $150000,Above $15000
Dwelling Type	0,1	Rent,Own

Gender	0,1	Female,Male
Marital Status	0,1	Single,Married or Cohabitation

Nouvelle Restaurant: is an updated version of L'Experience Restaurant (see below) and used as a case study dataset in ***Marketing Research,*** 10th edition, by Burns and Veeck.

Friendly Market: this dataset pertains to a single case study at the end of Chapter 14 in ***Basic Marketing Research*** that requires the use of crosstabulation. It pertains to a mom-and-pop convenience store using friendly customer service to compete against a neighboring national chain convenience store. It has mostly categorical variables.

L'Experience Restaurant: this is a simulated survey dataset for a restaurant concept being tested in a major metropolitan area. It is the most recent variation for this case study found in ***Marketing Research***, 9th edition, by Alvin Burns and Ann Veeck (Pearson, 2020). Respondents have given their reactions, with respect to desirability, of several possible features of this new restaurant. Demographic and media usage variables are included.

Pew American Trends: Pew Research conducts public opinion research studies on a regular basis, and one of its main data sources is its American Trends Panel Survey conducted quarterly. According to its website, "Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping the world. It conducts public opinion polling, demographic research, media content analysis and other empirical social science research. Pew Research Center does not take policy positions." The organization releases these survey databases to the public approximately 2 years after they take place. The dataset for the survey conducted in 2017 has been organized into an XLDA dataset.

Financial Well Being Survey: Conducted for the Consumer Financial Protection Bureau in 2016, this extensive survey involves Americans' abilities to: control finances, absorb financial shocks, meet financial goals, and enjoyment of life. The survey is the basis for the Bureau's report, ***Financial Well-Being in America*** (2017).

Medical Conditions 20YearsOldandOlder: The Centers for Disease Control conducts an annual survey covering a wide variety of medical issues. This dataset is the result of merging the demographics dataset with the medical conditions dataset. Most of the medical conditions questions are asked only of respondents 20 years of age or older, so respondents younger than 20 years old are deleted from the dataset. The most recent data release is 2015 and is used in this dataset.

PEW COVID Survey. See the Pew The American Trends Panel (ATP) above. Data in this survey is drawn from the panel wave conducted April 20 to April 26, 2020. A total of 10,139 panelists responded. The survey pertains to attitudes, opinions, and behaviors related to COVID-19.

Mobile Technology and Home Broadband. Dataset for survey on use of cell phones, smart phones, broadband internet, social media, etc. conducted by Pew Research in 2019 with U.S. representative sample of about 1,500 respondents.

Why Americans Don't Vote. 2020 online survey dataset identifying voter types (Rarely/Never, Sporadic, Always) with questions on beliefs, trust in officials, personal impact, etc. conducted by Ipsos for FiveThirtyEight. The dataset has over 5,800 respondents and is made available on GitHub.

GSS(General Social Survey). Huge dataset from a 2018 U.S. survey on a myriad of social issues, beliefs, opinions, and behaviors that is conducted every 2 years in affiliation with NORC at the University of Chicago. The survey dataset, available to any interested user, has over 1,000 questions and more than 2,300 respondents.

Other Datasets: There are no plans to expand these datasets; however, if an especially interesting or informative dataset is identified, it may be added.

Basic Data Analysis Concepts

Operation of XL Data Analyst Pro is simple, but users should be knowledgeable of basic data analysis concepts in order to fully utilize it. Following are descriptions of data analysis concepts that are relied upon consistently in this manual. Each concept is defined in nontechnical terms. In isolated cases, concepts that pertain to a specific data analysis type are described in its procedure description found in subsequent sections of this manual.

Dataset: using the notion of a survey, a data set is the resulting raw data organized in columns and rows (such as a spreadsheet). The columns pertain to questions or parts of questions in the survey, while the rows correspond to respondents or cases. For example, column 1 contains the answers to question #1. Alternatively, a dataset is any set of numbers and/or text (although text is not recommended for XL Analyst Pro) arranged in a rows-by-columns matrix.

Variables: a variable comprised of data in a column in a dataset, and it usually pertains to a question or a question part in a survey. A variable is usually identified with a descriptive name or label such as "Satisfaction with the experience," or "Income category."

Data Codes and Data Labels: raw data in a dataset may be of various forms. One form is code numbers that stand for answer labels to a particular question on a survey questionnaire. For, example, with a "yes" or "no" type question, the data codes could be "1" for "yes" or "0" for 'no." With a multichotomous question (one with more than 2 answers such as what is the color of one's automobile), it is customary to code the answers, "1," "2," "3," and so on to account for each of the colors in the order in which they are listed as eligible answers.

A special-purpose code is a **Midpoint Code** where the analyst has calculated the midpoint of an interval such as "25" for the midpoint between age "20 to 30" which is one of the ranges appearing on a question about a respondent's age. Such ranges are often used on questionnaires for streamlining and ease of response purposes. It is acceptable to have a midpoint code that contains decimal places such as "37.5." Using the midpoints rather than arbitrary codes of 1, 2, 3, etc. transforms the coding system to approximate numbers that are faithful to the underlying scale. In this way, a calculation such as the average age, results in a value that approximates the average age if each respondent indicated his or her precise age.

Another special-purpose code is an **Equivalency Code** where the analyst has specified values or magnitudes that effectively correspond to a variable's levels or categories. For instance, a high school diploma is typically earned in 12 years while a bachelor's degree is usually earned in 16 years, so number of years may serve as codes for educational attainment levels.

Again, users may be analyzing datasets that are not survey-based, in which case the coding systems for the variables are those established by the originators of the datasets.

Categorical Variable: a categorical variable is one where the answers do not have any magnitude differences. Also known as nominal data, a categorical variable's possible answers are simple categories or classes such as "blue," "red," or "yellow." Because there is no magnitude difference between the categories, only certain analyses are appropriate. Normally, categorical variables are analyzed by simple counts such as the number of times each response is recorded. These counts can be converted to percentages.

Metric Variable: On the other hand, a metric variable is one where there are known magnitude differences between the possible answers. Examples are the number of years of one's age, the

number of dollars spent purchasing an item, or the number of miles driven to the nearest fast-food restaurant. Sometimes called interval or ratio or cardinal measures, metric variables allow for "higher level" analyses using division such as averaging or computing a standard deviation.

With survey data, it is sometimes useful to identify a natural metric variable versus a synthetic metric variable. A **Natural Metric Variable** is one that uses a natural numbering system such as years of age, dollars of expenditure, miles driven, times frequented, or the like. A **Synthetic Metric Variable** is derived from a scale designed to capture emotion, attitude, opinion, agreement, satisfaction, or some other emotive measurement. Synthetic metric scales are sometime called "intensity" scales as they measure the intensity of a feeling or opinion usually ranging from strong negative to strong positive. Synthetic metric scales are constrained to usually 10 or fewer units such as 1-10, 1-7, or 1-5 scale, while natural metric scales and not constrained.

Data Analysis: Data analysis is a computational procedure applied to specific variables (i.e. columns of data) in a dataset. Analysis is performed on variables using their raw data numbers and/or data codes. As is the case with practically all data analysis programs, XL Data Analyst Pro will perform whatever analysis is chosen on whatever data is specified by the user. Consequently, it is incumbent on the user to be aware as to what variables are metric and which ones are categorical and to specify them correctly when choosing variables for analysis. *Categorical-versus-metric variable requirements are described in subsequent descriptions of the various XL Data Analyst Pro analyses.*

Outliers: values in a dataset can be extreme, and it is useful to identify and possibly remove very high or very low values in a variable's data. The most common outlier identification method is the use of ±3 standard deviations from the average. For example, with an average of 10 and a standard deviation of 1, values below 7 (10-3x1) and above 13 (10 + 3x1) would be outliers. With normally distributed data, values outside ±3 standard deviations boundaries theoretically have a less than 1% probability of representing the dataset. With XL Data Analyst Pro, users may opt to have outliers removed from the data prior to each analysis to eliminate the influence of extremes on the findings.

An alternative outlier system is the use of quartiles, specifically the interquartile range (IQR) or range from the 1st to the 3rd quartile values. 1.5 times this value is applied with low outliers being data below the median minus1.5* IQR and high outliers being above the median plus 1.5*IQR. This system is sometimes referred to as "fences."

XL Data Analyst Pro uses ±3 standard deviations as the default and allows for ±1.5 IQR if the user desires. The numbers of high and low outliers are always reported, and the user may opt to have them removed.

Statistical Significance/Level of Confidence: traditional (classical) statistical analysis procedures are based on certain assumptions or beliefs that permit statements as to the statistical significance of, or confidence in, the findings. Because the 95% level of confidence is almost universally used in practical situations, meaning analyses by marketing researchers, political analysts, public opinion pollsters, education researchers, and managers in general, XL Data Analyst Pro's default is the 95% level of statistical confidence.

In other words, except for the simplest analysis (termed "Summarize" in the case of XL Data Analyst Pro), some appropriate statistical test is applied to the finding, and the user is assured that the finding holds for 95% of the time. A convenient notion is that if the research were replicated (identical sampling, identical questions, same sample size, etc.) the analyst is 95%

confident that the same finding would result. Alternatively, if 100 replications took place, at least 95 of them would result in the current finding. (Typically, "finding" is a range rather than an exact value.)

However, researchers differ in their risk acceptance, and some prefer to use 99% level of confidence (risk averse) or 90% (risk prone). *Consequently, with XL Data Analyst Pro, the user may specify any significance level between 1% and 99%.* The specified level of confidence is identified by XL Data Analyst Pro for any analysis utilizing statistical significance. In addition, the computed statistical values such as F, t, or z are provided. However, this statistical information is not a conspicuous aspect of XL Data Analyst Pro output.

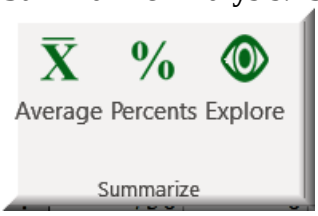
Effect Size: statistical significance is largely affected by sample size, and researchers often work with large samples. It is useful to interpret statistical significance with effect size which is a measure of the magnitude of evidence for the finding. For example, a correlation may be found to be statistically significant, meaning different from 0, but its absolute size is its effect size. Thus, a significant correlation of .8 is considered "strong;" whereas a significant correlation of .2 is "weak." Effect size measures vary by type of analysis, but with most analyses, the measure is commonly labeled as "large," "medium," "small," or "very small." Thus, statistically significant findings that are small or very small in effect size are usually not meaningful. *XL Data Analyst Pro always reports effect size labels for statistically significant results.*

Data Visualization: Recently, researchers have adopted the use of visualizations such as graphs and infographics to better communicate findings to clients or others who do not relate to tables or statistics. Data visualization has many benefits including: faster comprehension, intuitive presentation, memorable impact, and overall professional appearance. *XL Data Analyst Pro has a variety of graph templates for the findings of its Percents module, and it creates graphs for any statistically significant findings identified in other modules.*

Overview of XL Data Analyst Pro Analyses & Utilities

This section has a quick description of each of the several different data analysis types contained in XL Data Analyst Pro. Detailed descriptions of each type of analysis, its menu procedure specifics, and resulting output tables are provided in the following section.

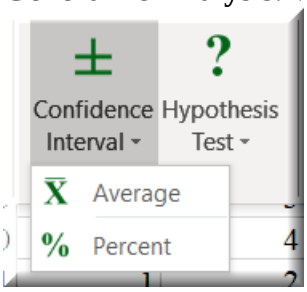
Summarize Analysis: Summarization analysis describes the typical response or typical respondent with the use of a mode or an *average*. The mode is the most frequently occurring value in a string of values, while the average is determined by summing up all values and dividing by the total number of times all values occur. Categorical variables should be analyzed with a mode; whereas metric variables are analyzed with an average or mean.



A separate concern with summarization analysis is the degree of agreement among the responses or numbers. With a categorical variable, a frequency distribution (number of times each code is mentioned) or *percent distribution* (percent of times out of 100% each code is mentioned) depicts the similarity or variability in the responses. With a metric variable, the standard deviation is normally computed to relate the variability in the data. The range (minimum and maximum values) is also sometimes used to assess variability with a metric variable.

Exploratory data analysis employs a host of measures and graphs that describe various characteristics of a variable's data, and advanced users may want to inspect them. Consequently, the "Explore" module is part of Xl Data Analyst Pro's summarization analyses.

Generalize Analysis: With a survey, individuals who answer the questions are referred to as a "sample" of respondents. This sample represents a population or all the individuals from which the sample is drawn. Classical statistics assumes that the sample is "random" or selected such that all members of the population are represented proportionately in it. Because random samples do not include all members of the population, generalization analyses of various types are used to make statements that accurately describe the population while taking into account that a sample is used.

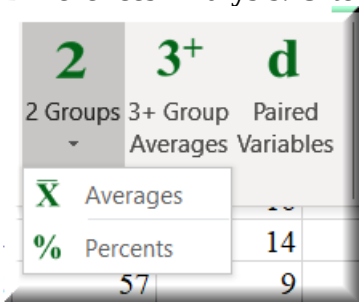


A *confidence interval* using a given level of confidence is applied using the variability found in the sample. With a metric variable, the standard deviation is used, and with a categorical variable, the standard error of the percent is used to calculate a lower and upper boundary for the confidence interval. Thus, the analyst can declare that the population value (either the mean or the percent) falls between the lower and upper boundary with given level of confidence that this is the case. As noted earlier, various interpretations of (e.g.) 95% confidence exist, but a convenient one is to state that if the survey were repeated independently with identical sample method and sample size 100 times (replications), then 95% of these samples would find their averages or percents to lie between the lower and upper boundaries of the confidence interval.

Another form of generalization analysis is a *hypothesis test* where the analyst (or perhaps a client) states a priori to the dataset's analysis that the population average or percentage occurrence of some category will be a certain value. Using hypothesis test equations, the analyst can assess the degree to which the random sample supports or fails to support this hypothesis. With the (e.g.) 95% level of confidence, the analyst will support the hypothesis if it is found to fall within the 95% confidence interval or fail to support it if it is found to fall outside the 95% confidence interval boundaries. However, with a hypothesis test, the confidence interval is not

computed as the analyst calculates a z value and compares it to ±1.96 (for 95% level of confidence) as an efficient means of determining support or nonsupport for the hypothesized value. Again, an XL Data Analyst Pro user can specify any level of confidence for any analysis, and the corresponding z value will be used.

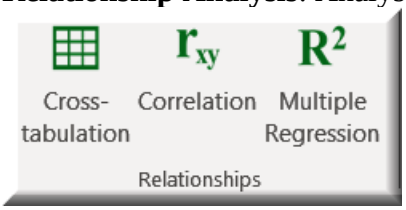
Differences Analysis: Often an analyst is interested in comparing groups within a population.



Subgroups such as males versus females, young versus old individuals, or people who dwell in residence types such as condominiums, apartments, mobile homes, and so on can be compared as to how different they are with respect to some variable of interest. Subgroups (categorical grouping variable) can be compared based on the average answer for a metric variable such as age or the percent responses for a categorical variable category such as percent "yes" answers to whether they use Uber.

Two group differences tests can be computed for either metric or categorical target variables, while *three-plus group differences* can be applied for metric target variables. When groups are found to be significantly different (at given level of confidence), there is enough evidence to conclude that the group differences are stable, meaning that the differences hold at the given level % of the time across replications of the sample survey, whereas, if not, these differences will not hold across replications. A separate *paired differences analysis* can be applied to determine if one metric variable's average differs significantly from another one's as for instance how many times on average people buy online versus how often they buy the same product in a retail store.

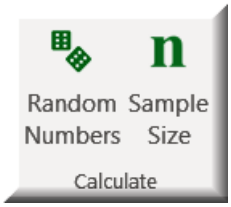
Relationship Analysis: Analysts are sometimes interested in the strength and direction of the



relationship of one variable to another or others. Such relationships, if found statistically significant at the (e.g.) 95% level of confidence, can be relied upon to hold across replicated samples or in the population, and such relationships can be relied upon for strategic recommendations. With two categorical variables, *cross-tabulation analysis* is applied, whereas, with two metric variables, *correlation analysis* is applied. With cross-tabulation, the analyst might determine that males are much more likely to purchase Nike brand shoes that are females, while, with correlation, one may be determine that younger individuals spend more money on food delivery than do older ones. In either case, the analyst can attest that the relationship is stable in the relevant population. Significant correlation also infers direction and strength of relationship.

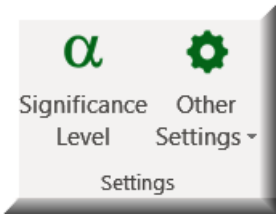
An entirely separate analysis is *multiple regression*, where a set of independent, mostly metric, variables is used to predict the level of a single metric dependent variable. With this analysis, the analyst identifies those independent variables that are (e.g.) 95% likely to be predictors of the dependent variable and determines the relative amount of this relationship with the use of a standardized coefficient for each significant independent variable. The strength or goodness of a multiple regression analysis finding is assessed by way of the amount of "explained variance."

Calculations. Researchers sometimes have need for specific calculations, and XL Data Analyst Pro provides two such procedures. With random numbers, it will provide a list or table of numbers that are randomly distributed across a specified range. At the user's option, these numbers may be unique or repeated. The other calculation is for sample size, and the XL Data Analyst uses the standard sample size formula to calculate sample size at a given desired accuracy. The output provides sensitivity analysis for various accuracy levels, and if the user provides cost information, sample cost estimates are calculated.



Utilities. Two utilities features are available to the use. With Clean-up, XL Data Analyst Pro will check the integrity of the user's data, variable names, value codes, and value labels. It also corrects any data errors (such as formulas or Excel errors) in the data. Data utilities include filtering or selecting subsets of data for analysis, unfiltering, importing data in comma separated variable (CSV), Excel (XLS), or XLDA format, and exporting the current XL Data Analyst Pro work into an Excel workbook file. (Note: import and export is not available on Macs.)

Settings. The significance level setting allows the user to specify a desired level of confidence for all analyses. The specific level of confidence is reported on all output that relies on statistical significance. Other settings include: identification with specified headers for the output, format of tables lines, tables color, graph types, specification of outlier method, restoration of all default settings, and variables inspection.



Descriptions of XL Data Analyst Pro Analyses

XL Data Analyst Pro performs several commonly used analyses. In the following descriptions of these analyses, a conscientious effort is made to not include complex formulas. Certain statistical concepts are referred to but not defined nor described using formulas. Users are advised to consult statistics references for technical definitions, explanations, and formulas.

Screenshots, some with annotations, are provided for all the following descriptions. In all cases, the Homespread Restaurant XLDA dataset is used (which is the dataset for the downloaded XL Data Analyst Pro .xlsm file.) After a brief description of the analysis, an annotated screenshot of the window used to select variables for processing the analysis is provided under "Procedure," while the resulting annotated analysis findings output screenshot is provided under "Discovery." In some cases, the exhibits have been modified or resized to eliminate the need for scrolling, to reduce empty worksheet space, or to better fit the margins of this manual.

Note: Some screenshots may not be precisely faithful to the current version of XL Data Analyst Pro as it is being constantly tweaked with improvements.

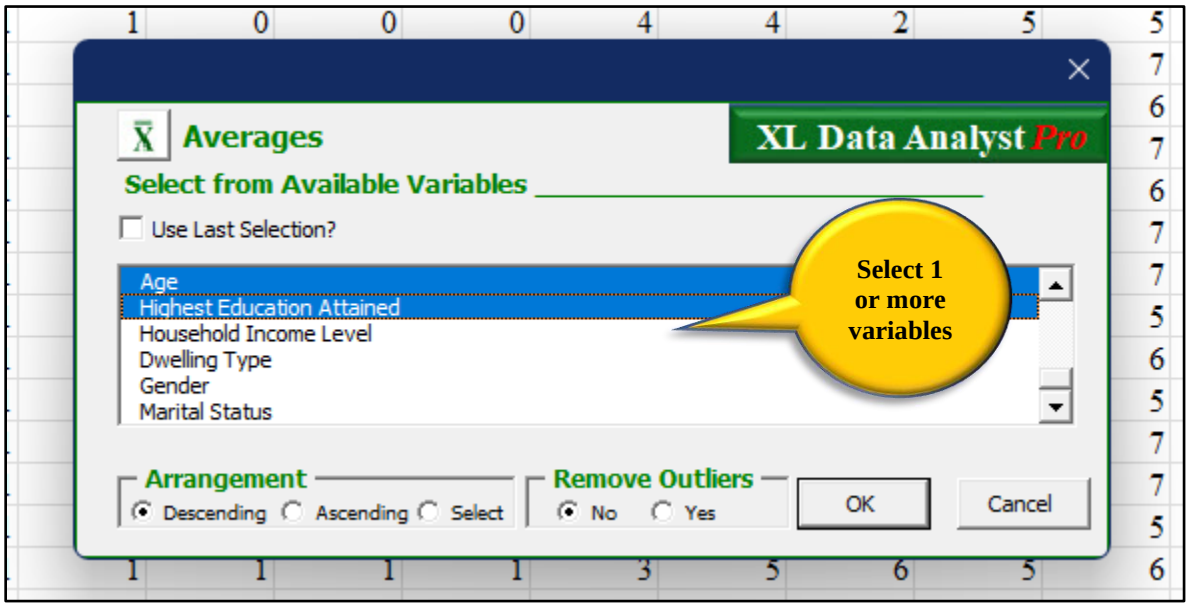
Summarize Analyses

Summarization analyses involve averages for metric variables and percentage distributions for categorical variables. Summarize analyses are sometimes called "descriptive analyses" as they describe the typical response for a particular question in a survey or typical value for a variable. In addition, summarization analyses provide information on the typicality or variability of the data as for example a percentage distribution of all values for a variable, the range (minimum and maximum values), or the standard deviation. In the case of percentage distributions, graphs such as pie charts are often used to provide a visual representation of the data.

Average Analysis

Averaging involves the summing of all values for a metric variable and dividing that sum by the number of occurrences of all values which is the effective sample size. Only numeric values are averaged: alphameric or other data (such as blanks) are ignored. The standard deviation is calculated, and the range (upper and lower values) is determined. The variable's Description is used to identify the variable in the resulting table.

Procedure: Use the Average icon which opens the Averages selection window. Select the variable(s) by highlighting in the "Select from Available Variables" window. Anywhere from one to all the variables can be selected.

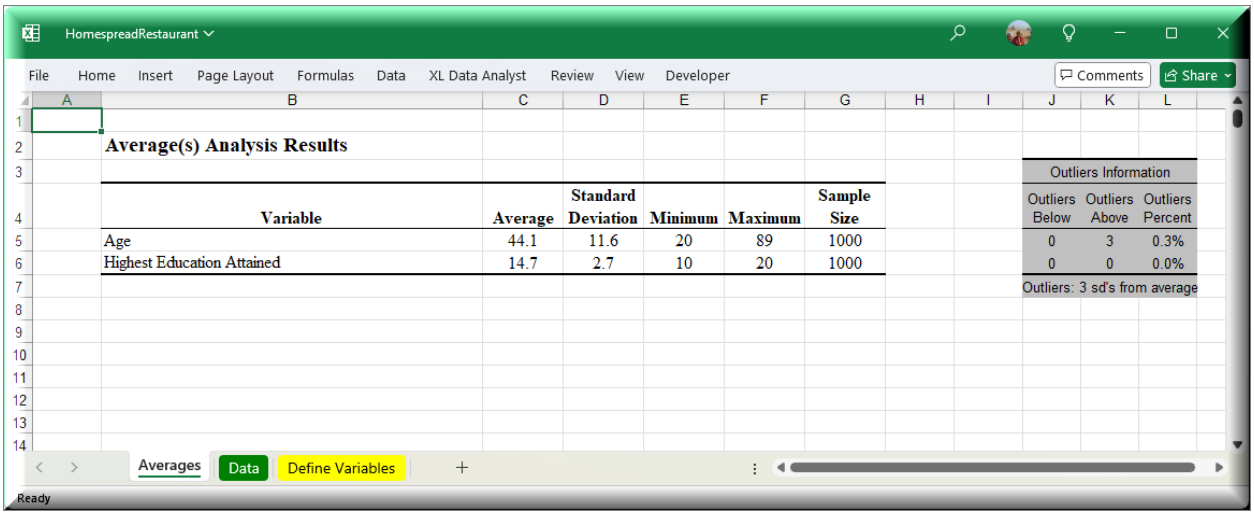


Averages Variable Selection Window

Options: There are three options available on the Averages window:

- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.
- Arrangement of the variables in the output table by their averages in descending order, ascending order, or the order in which they were selected. Default is "Descending."
- Remove Outliers? will find and remove outliers. Default is "No."

Discovery: On an "Averages#" output worksheet, each analyzed variable is presented in a table with its average, standard deviation, minimum value, maximum value, and sample size. If more than one variable is selected, the table will list the variables in the desired order of the computed averages. Outliers information is provided in a separate, parallel table. If outliers removal is selected, a note to this effect will appear under the averages table.

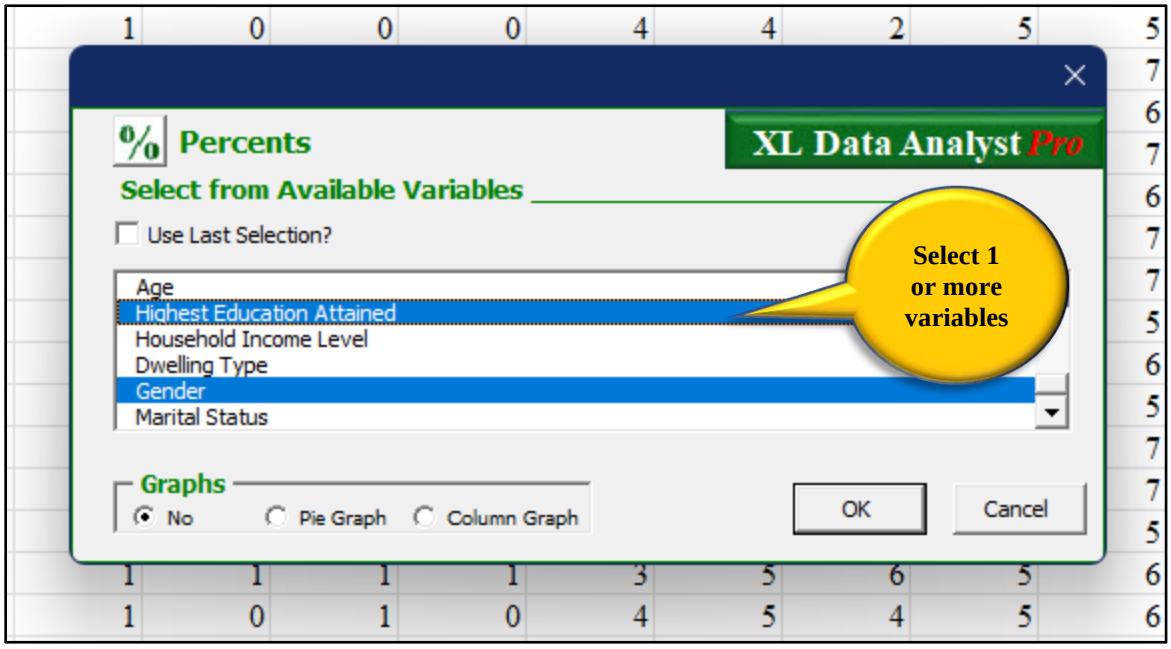


Averages Discovery

Percents Analysis

With percents analysis, all possible data code values are identified, and the number of occurrences, or count of each is determined. Each count is divided by the total number of occurrences, or sample size, to yield its percent. Percents analysis should be performed on categorical variables, but it can be performed with metric variables. The result is a frequency distribution of the counts of each value, the sum of the counts, and a percent distribution across the values. Thus, a frequency distribution and its accompanying percentage distribution table is provided as the result of the percents analysis procedure. The variable's data labels are used in its output table, and the Variable Description is used as the table's title. Any data code that is not specified on the Define Variables worksheet under the analyzed variable will appear in the table as the data code.

Procedure: Click the Percents icon to open up the Percents selection window. Select variables by highlighting them in the "Select from Available Variables" window. Anywhere from one to all the variables can be selected. The default is "No graph," and the user can select either pie or cylinder graph. Whichever graph type is selected is applied to all variables selected for analysis. In addition, the user may opt for XLDA Pro 3-D, XLDA Pro 2-D, XLDA Basic 3-D, XLDA Basic 2-D, or Excel 2-D graphs using the "Graph Format" feature in the "Format, Etc." Settings icon.

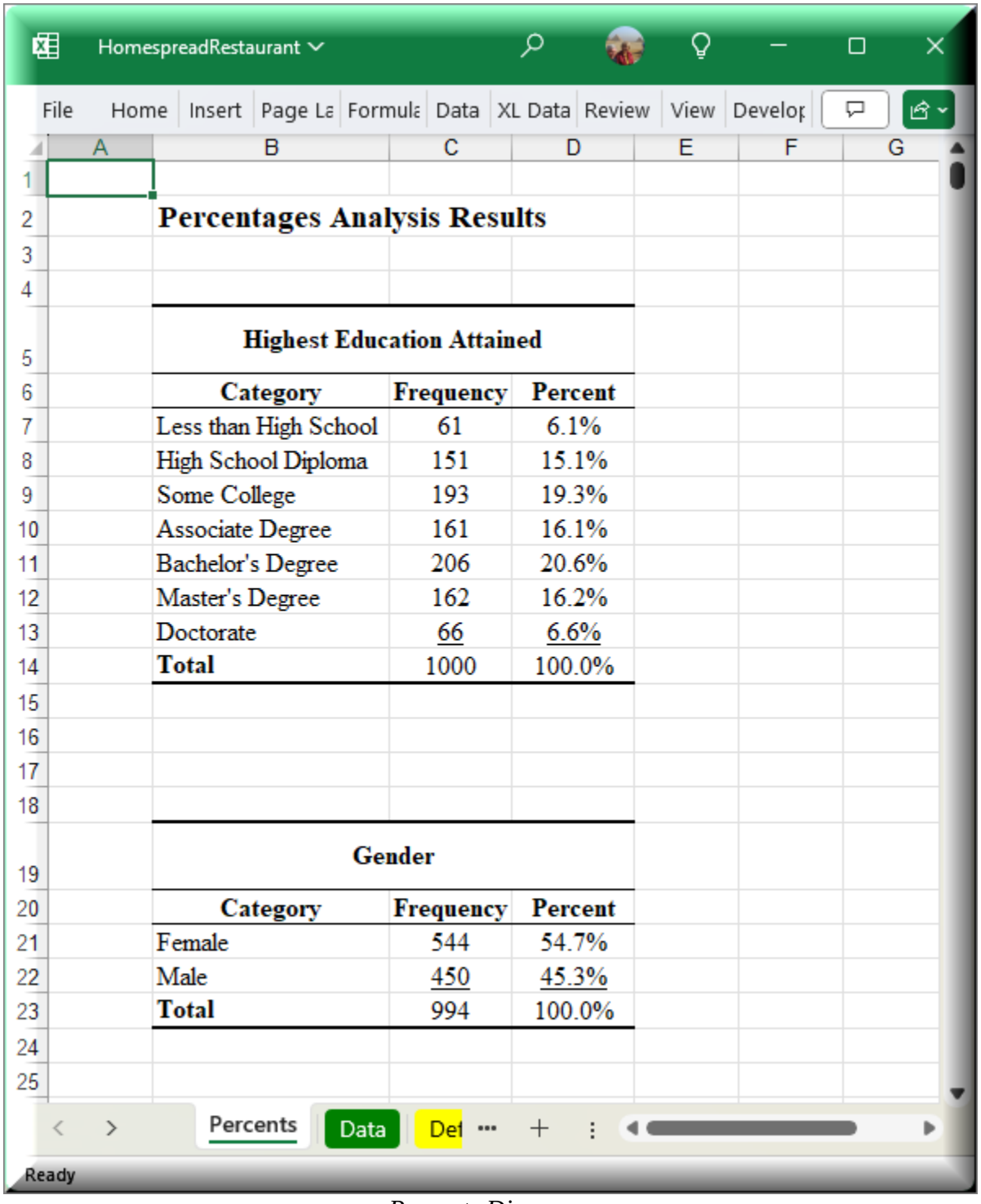


Percents Variable Selection Window

Options: There are two options available on the Percents window:

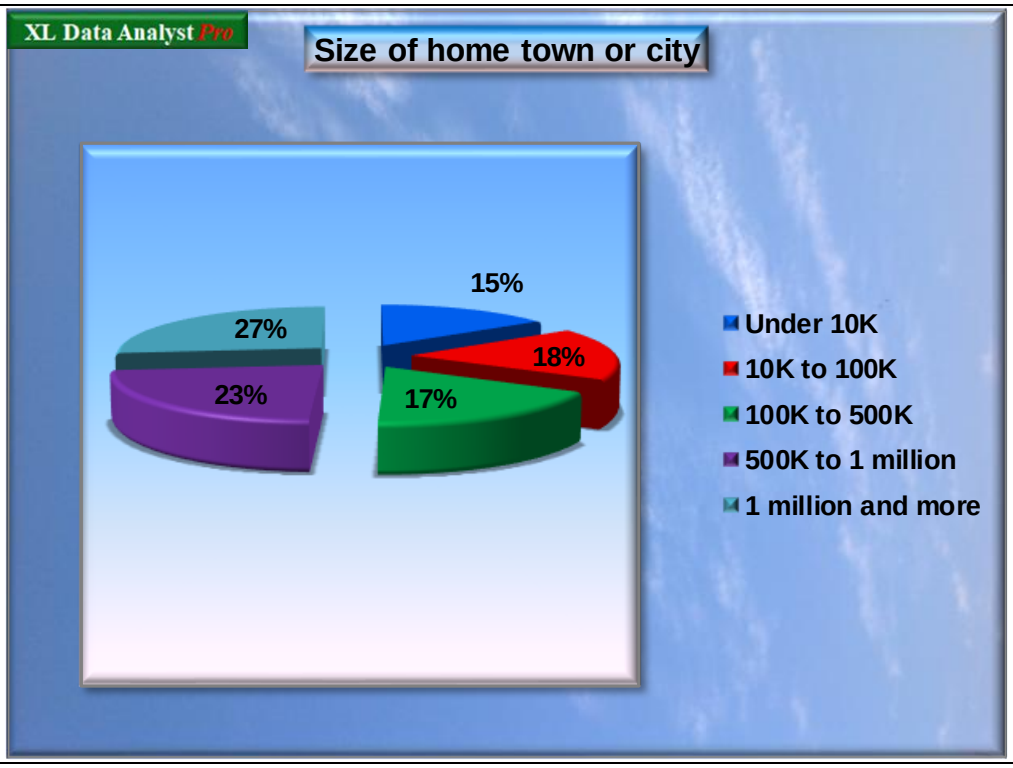
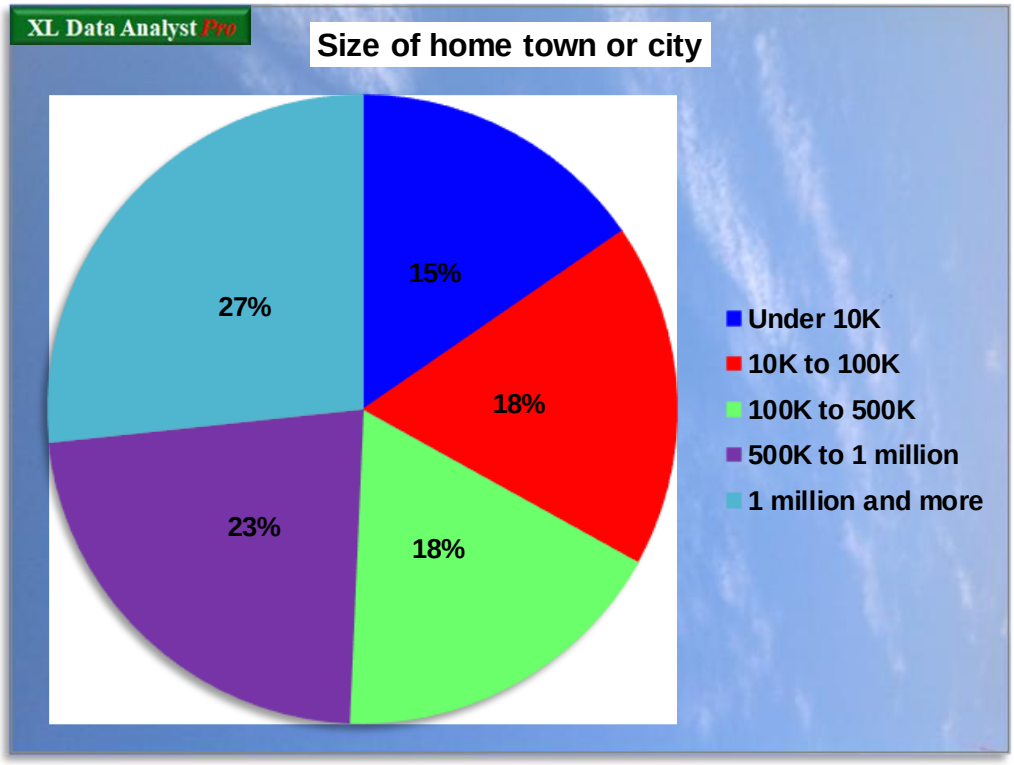
- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.
- Graphs – The user may select no, pie, or column graph. The default is "No Graph"

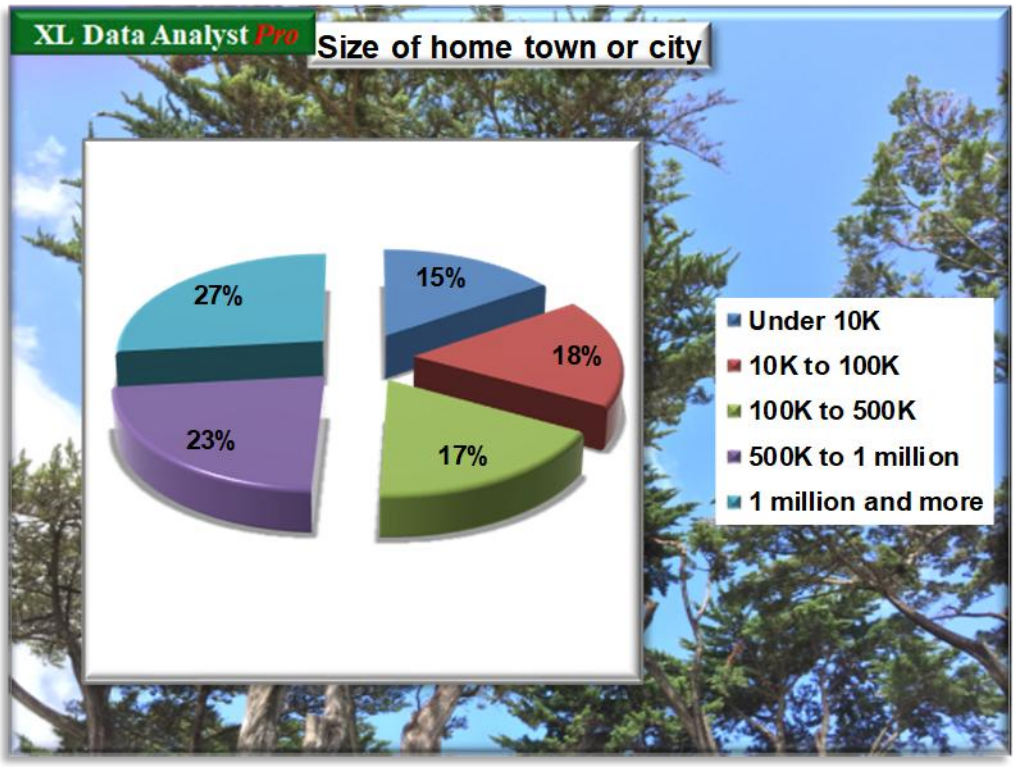
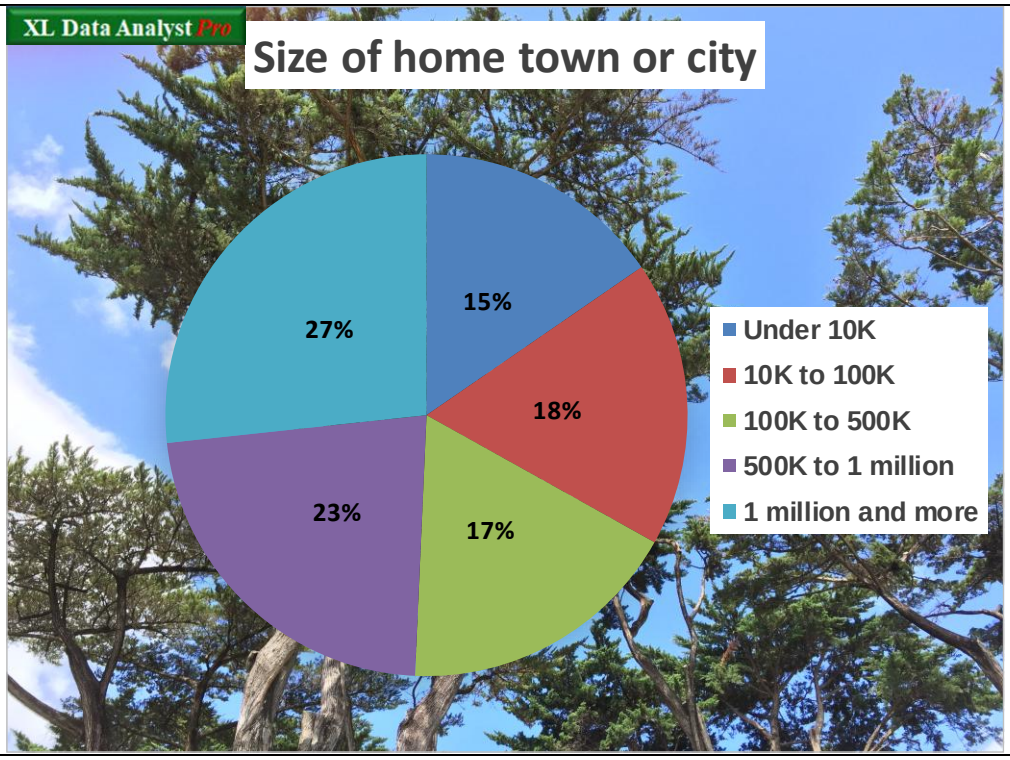
Discovery: On a "Percents#" worksheet, a percentage distribution table with accompanying frequencies is created for each variable selected. If selected, the appropriate graph will appear before each percentage distribution table. If more than one variable is selected, percentage distribution tables are created in the order of the variables selected.

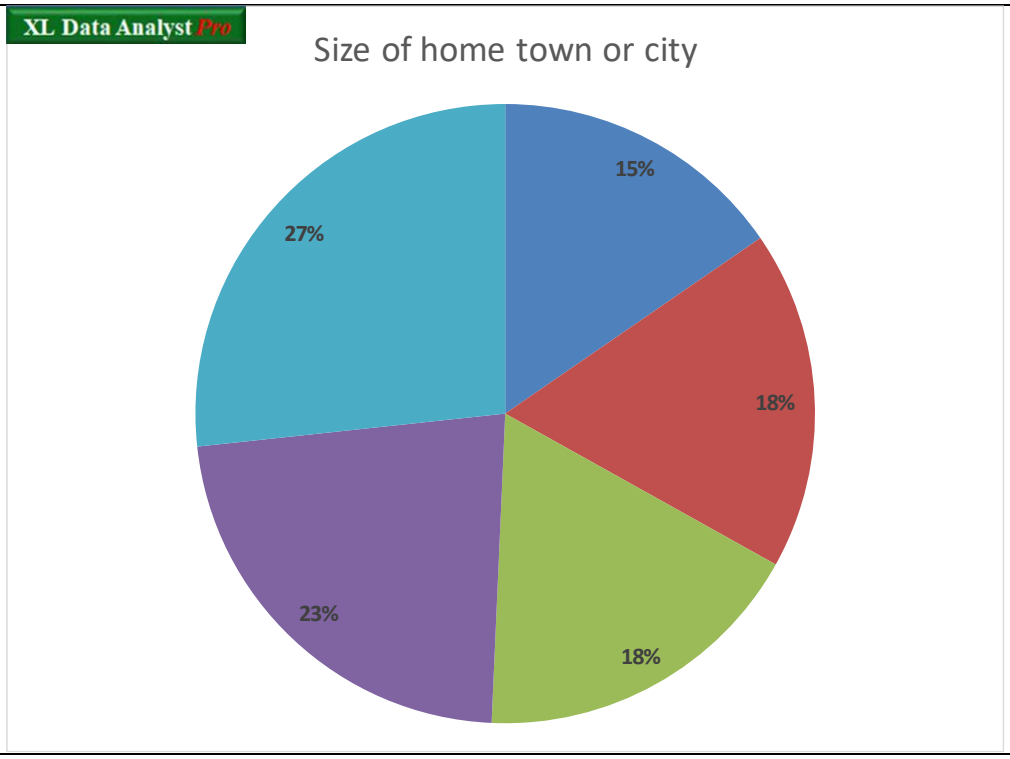


Percents Discovery

Percents Graphs: There are five graph format options with the Percents procedure. With whichever one is selected, if the pie or column graphs option is selected, XL Data Analyst Pro applies its template to the graph in the Percents module. The formats are noted below.

Graph Format	Pie Graph Template
Pro 3-D	
Pro 2-D	

Basic 3-D	
Basic 2-D	

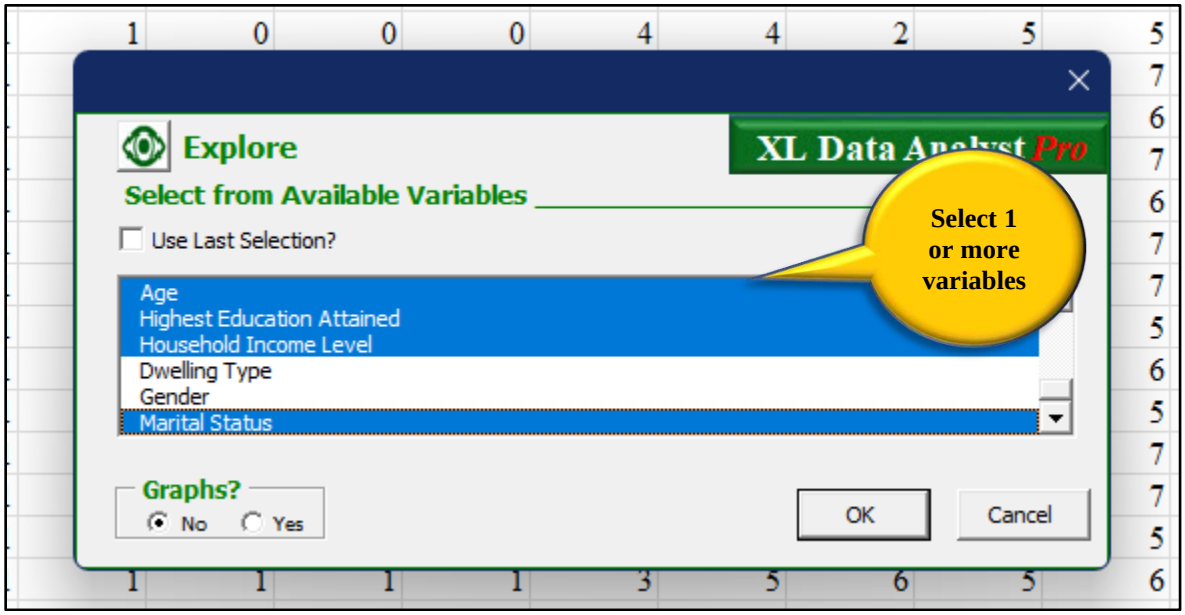
Excel 2-D	

XLDA Graphs Templates

Explore Analysis

Exploratory data analysis involves the use of a variety of measures for central tendency, variation, data distribution, and data quality. In addition, histograms, box-and-whisker, and Q–Q (quantile-quantile) plot (sometimes called normal probability plot) are used to examine a variable's data. Researchers use exploratory data analysis to become familiar with, summarize, and visualize a dataset.

Procedure: Click the Explore icon to open up the Explore selection window. Select variables by highlighting them in the "Select from Available Variables" window. Anywhere from one to all the variables can be selected.



Explore Variable Selection Window

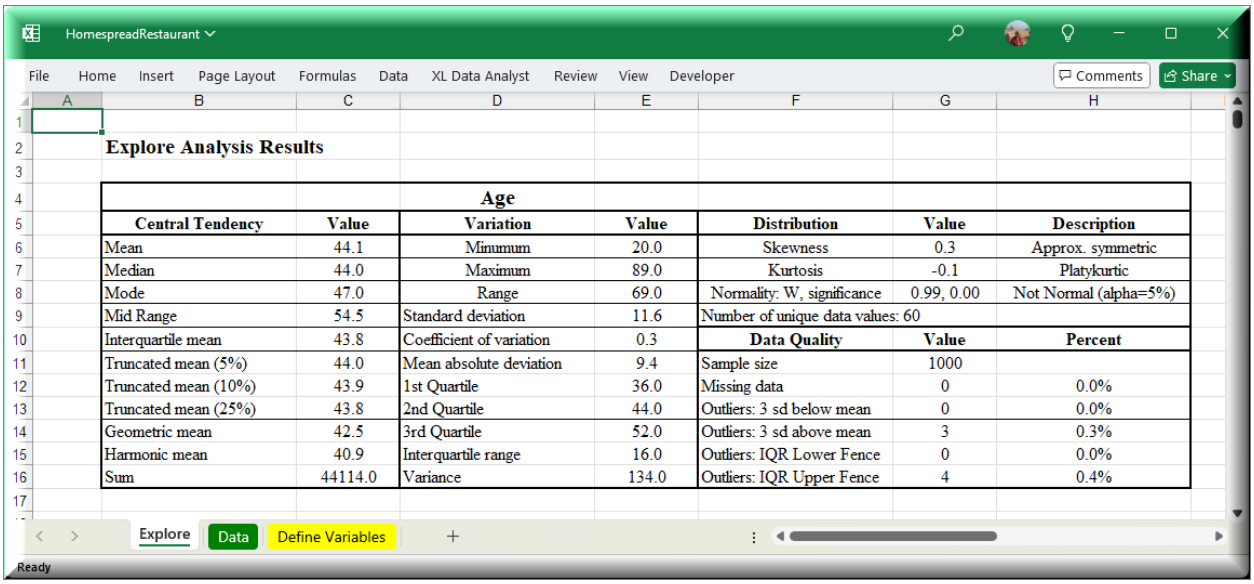
Options: There are two options available on the Percents window:

- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.
- Graphs? – The user may opt for histogram, box-and-whisker, and normal probability plot as part of each variable's discovery. The default is "No."

Discovery: On an "Explore#" worksheet, there are several of measures for:

- Central tendency
  - Mean
  - Median
  - Mode
  - Mid Range
  - Interquartile mean
  - Truncated mean (5%)
  - Truncated mean (10%)
  - Truncated mean (25%)
  - Geometric mean
  - Harmonic mean, and
  - Sum

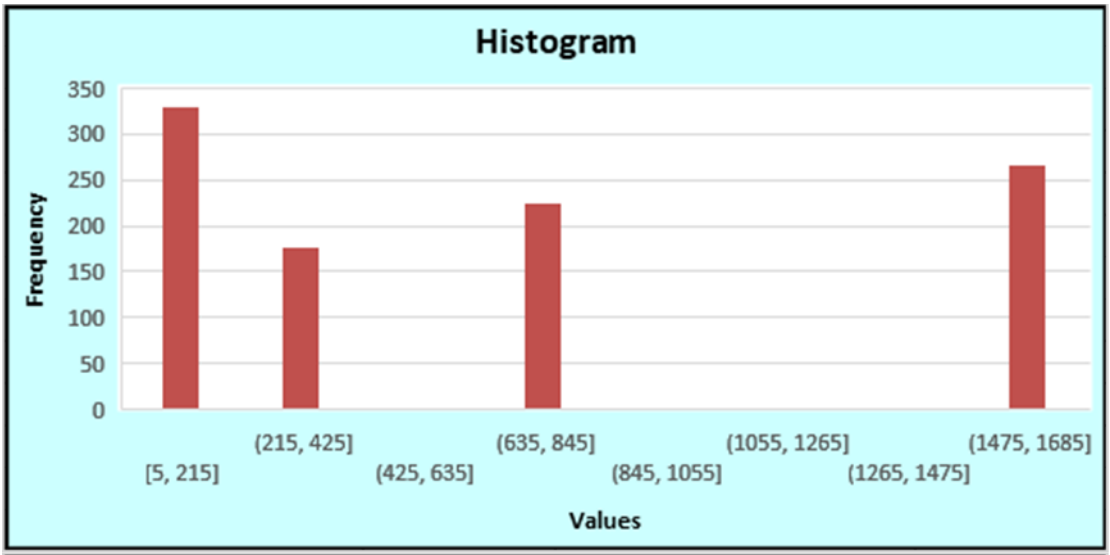
- Variation:
  - Minimum
  - Maximum
  - Range
  - Standard deviation
  - Coefficient of variation
  - Mean absolute deviation
  - 1st Quartile
  - 2nd Quartile
  - 3rd Quartile
  - Interquartile range, and
  - Variance
- Distribution
  - Skewness (value and description)
  - Kurtosis (value and description)
  - Normality (value and description)
  - Number of unique data values
- Data quality:
  - Sample size
  - Missing data
  - Outliers: 3 standard deviations below the mean (number and percent)
  - Outliers: 3 standard deviations above the mean (number and percent)
  - Outliers: 1.5 Interquartile range lower fence (number and percent)
  - Outliers: 1.5 Interquartile range upper fence (number and percent).

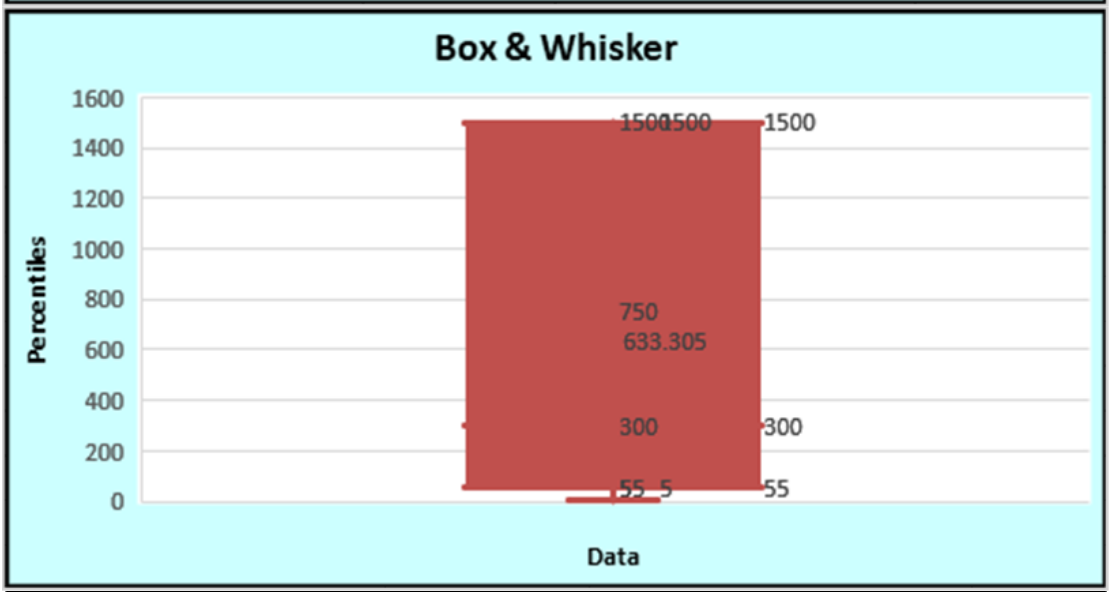


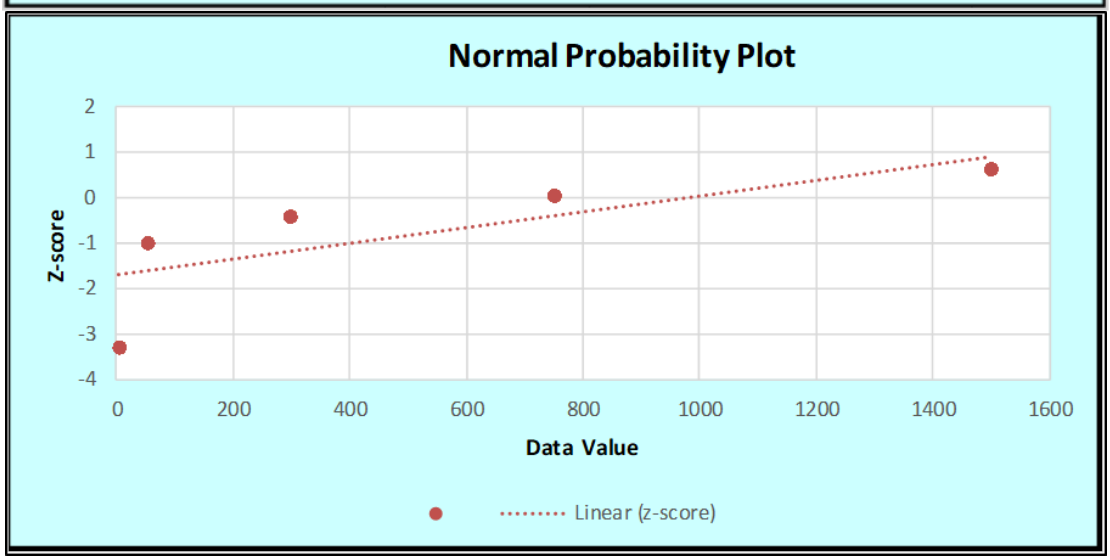
Age						
Central Tendency	**Value**	**Variation**	**Value**	**Distribution**	**Value**	**Description**
Mean	44.1	Minumum	20.0	Skewness	0.3	Approx. symmetric
Median	44.0	Maximum	89.0	Kurtosis	-0.1	Platykurtic
Mode	47.0	Range	69.0	Normality: W, significance	0.99, 0.00	Not Normal (alpha=5%)
Mid Range	54.5	Standard deviation	11.6	Number of unique data values: 60		
Interquartile mean	43.8	Coefficient of variation	0.3	**Data Quality**	**Value**	**Percent**
Truncated mean (5%)	44.0	Mean absolute deviation	9.4	Sample size	1000	
Truncated mean (10%)	43.9	1st Quartile	36.0	Missing data	0	0.0%
Truncated mean (25%)	43.8	2nd Quartile	44.0	Outliers: 3 sd below mean	0	0.0%
Geometric mean	42.5	3rd Quartile	52.0	Outliers: 3 sd above mean	3	0.3%
Harmonic mean	40.9	Interquartile range	16.0	Outliers: IQR Lower Fence	0	0.0%
Sum	44114.0	Variance	134.0	Outliers: IQR Upper Fence	4	0.4%

Explore Discovery

In addition to these measures, the user may opt for the three following exploratory analysis graphs.







Explore Graphs

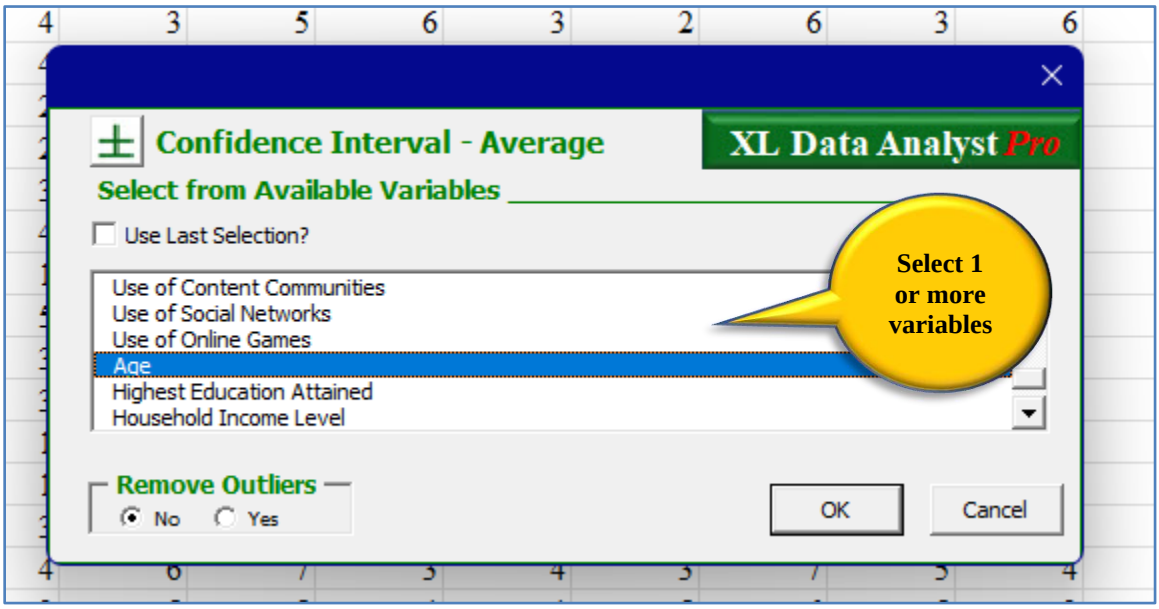
Generalization Analyses

Generalization analyses pertain to confidence intervals and hypothesis tests for percents or averages. As noted earlier, datasets are often samples representing much larger populations, and there are accepted classical statistical procedures that allow analysts to infer or generalize about the population. Assuming a random sample is used, due to sample size error and variability in the data, one type of generalization analysis, namely confidence intervals, results in a range - minimum and maximum - that the analyst is confident describes the population value (percent or average), while another type of generalization assess the degree of support for an a priori belief, called a "hypothesis," as to the population's value. As noted earlier, XL Data Analyst Pro utilizes a 95% level of confidence as its default, but the user may specify any desired level of confidence with the "Significance Level" feature.

Confidence Interval for an Average

The confidence interval for an average of a variable is computed via formulas that make use of the sample average, sample size, standard deviation, and standard error of the average. The result is a range, comprised of a lower boundary and an upper boundary, which defines the (e.g.) 95% confidence interval for the sample average. Users are (e.g.) 95% confident that this range accurately describes the population average.

Procedure: Use the Confidence Interval-Average icon sequence, and the Confidence Interval – Average window will open. Select one or more metric variables from the "Select from Available Variables."

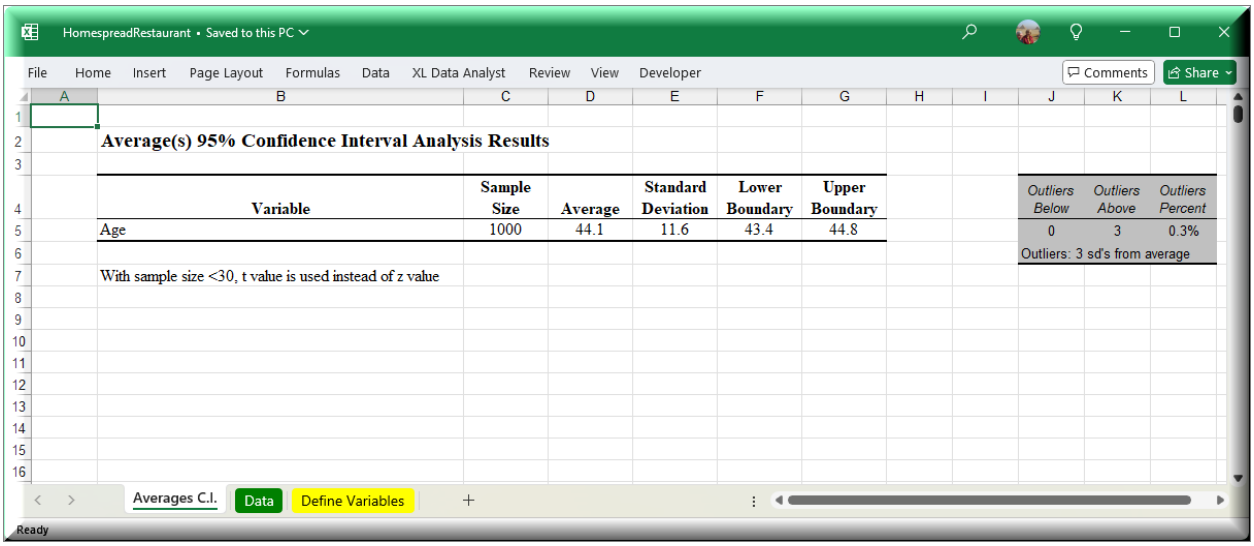


Average Confidence Interval Window

Options: There are two options available on the Confidence Interval -Average window:

- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.
- Remove Outliers? – The user may opt to have any outliers removed from the analysis if found. The default is "No"

Discovery: On an "Averages CI#" worksheet is table with: the sample size, average, standard deviation, and lower and upper specified level of confidence boundaries for each variable, listed in order of selection. Outliers information is provided in a separate, parallel table. If outliers removal is selected, a note to this effect will appear under the confidence interval table.

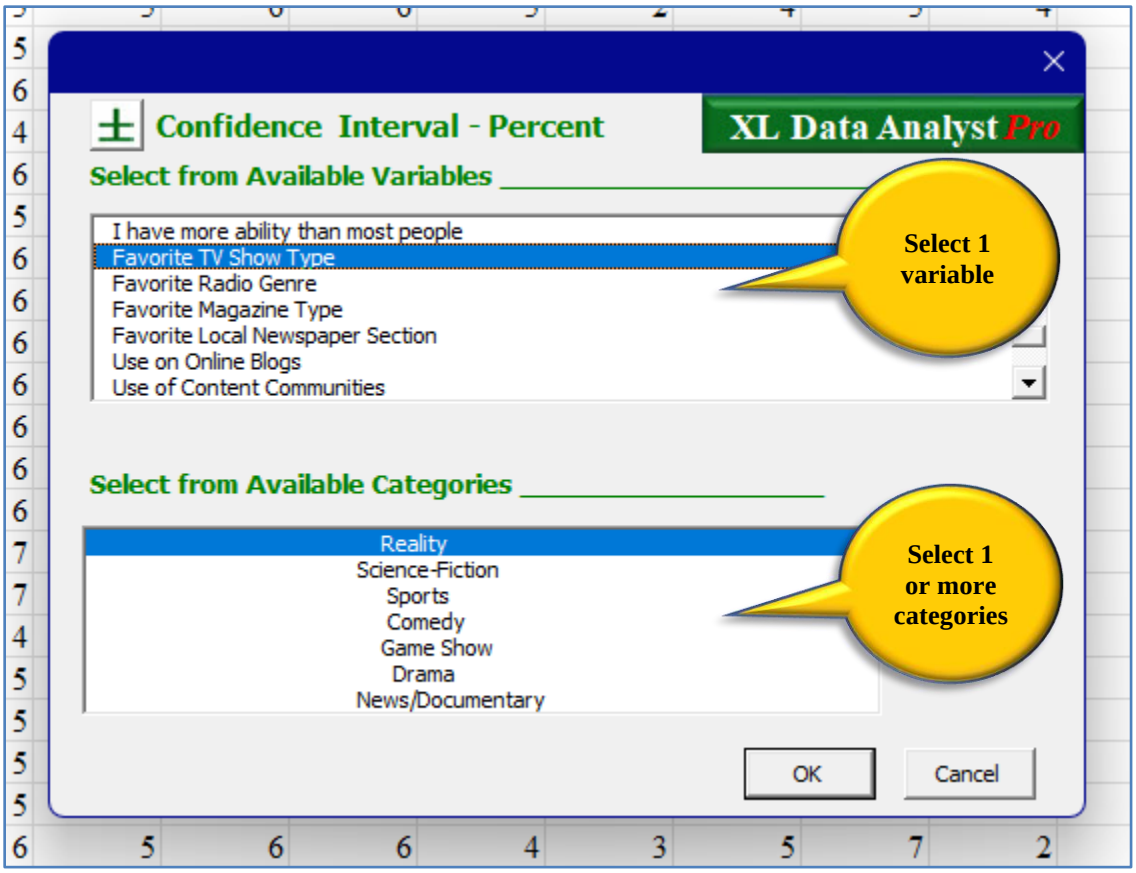


Average Confidence Interval Discovery

Confidence Interval for a Percentage

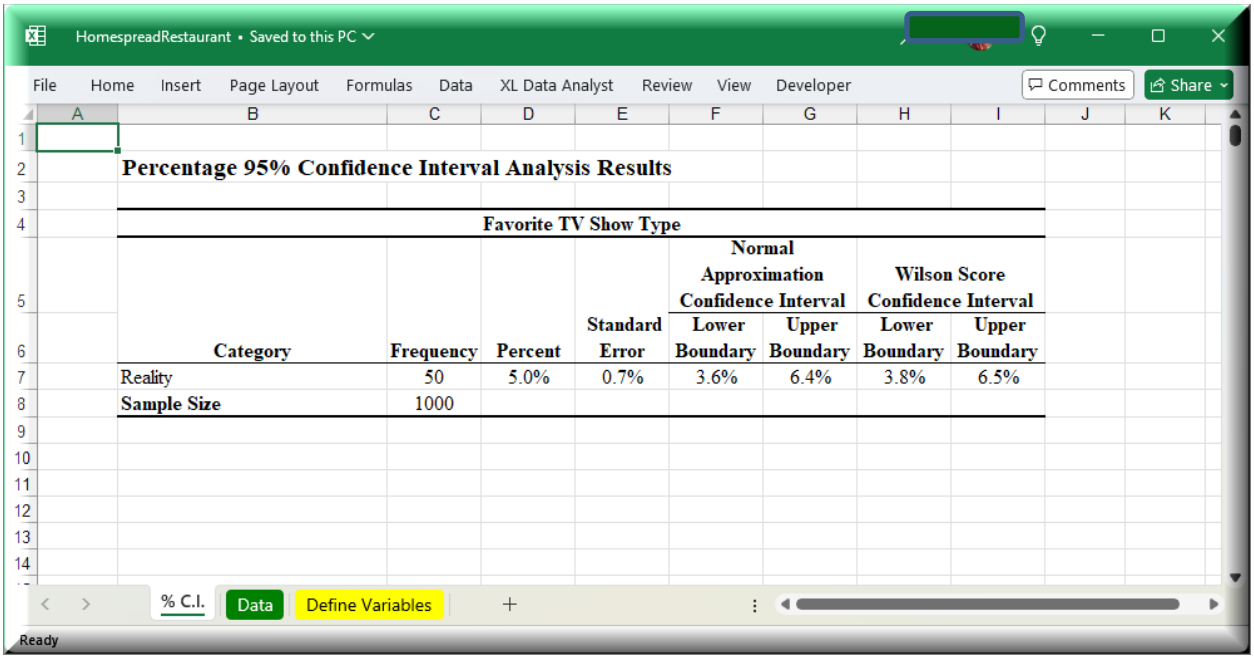
The confidence interval for the percent found for a particular category for one variable is computed via formulas that make use of the sample percent, sample size, and standard error of the percent. The result is a range, described as the lower boundary and the upper boundary, which defines the (e.g.) 95% confidence interval for the sample percent. Users are (e.g.) 95% confident that this range accurately describes the population percent.

±% **Procedure:** Use the Confidence Interval-Percent icon sequence. Select one categorical variable by highlighting it in the "Select from Available Variables" window. Select anywhere from one to all the selected variable's value labels that appear in the "Select from Available Categories" window.



Percentage Confidence Interval Menu Window

Discovery: On a "% CI#" worksheet, XL Data Analyst Pro provides a table with the variable Description as the table heading. Each of the selected categories is listed in order of selection with its frequency, sample percent, and standard error of the percent. Two types of confidence intervals are provided: Normal Approximation and Wilson Score. The Normal Approximation method is the standard method found in basic research methods textbooks, while the Wilson Score method is claimed to be an improvement especially in the case of small sample size and/or skewed data. It both cases, the table displays the lower boundary and upper boundary of its specified % confidence interval for that sample percent. The sample size of the variable is also indicated in the table.

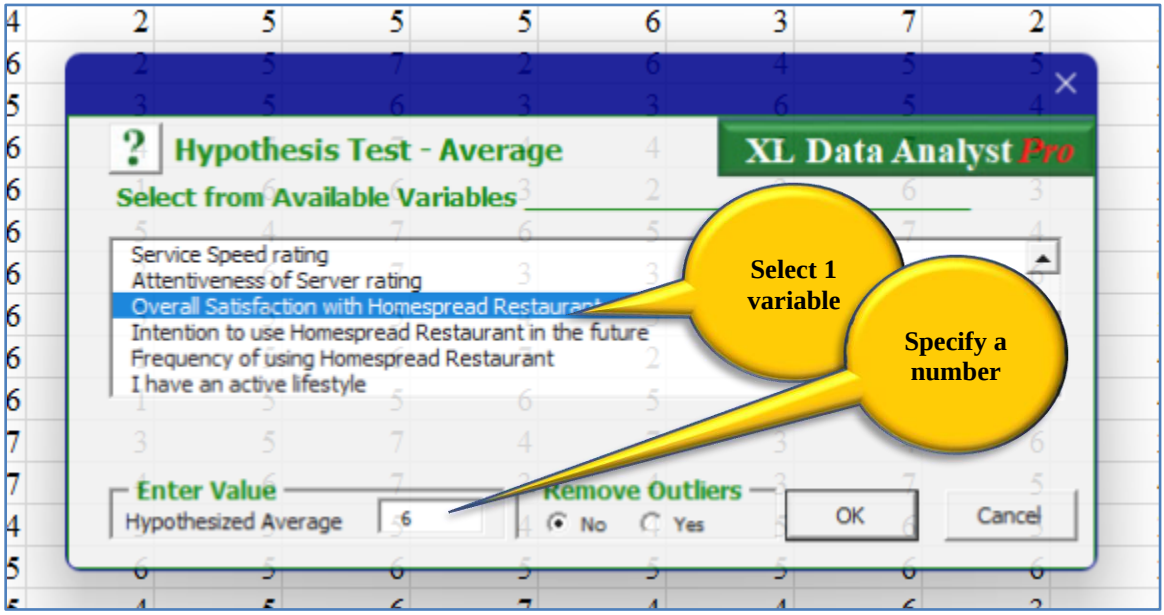


Percentage Confidence Interval Discovery

Hypothesis Test for an Average

With a hypothesis test, the analyst uses a random sample's findings to "support" or "not support" a belief as to the population average (mean) for the values pertaining to a metric variable. With an average hypothesis test, the sample average, sample size, standard deviation, standard error of the mean, and z value are used. There is a statement that the analyst is (e.g.) 95% confident of finding or no support for the hypothesis. The user may run this analysis with any desired level of confidence by using the "Significance Level" feature.

Procedure: Use the Hypothesis Test-Average icon sequence to open up the Hypothesis Test – Average selection window. Highlight to select a metric variable from the "Select from Available Variables" window and specify the hypothesized average in the "Hypothesized Average" input box.



Average Hypothesis Test Window

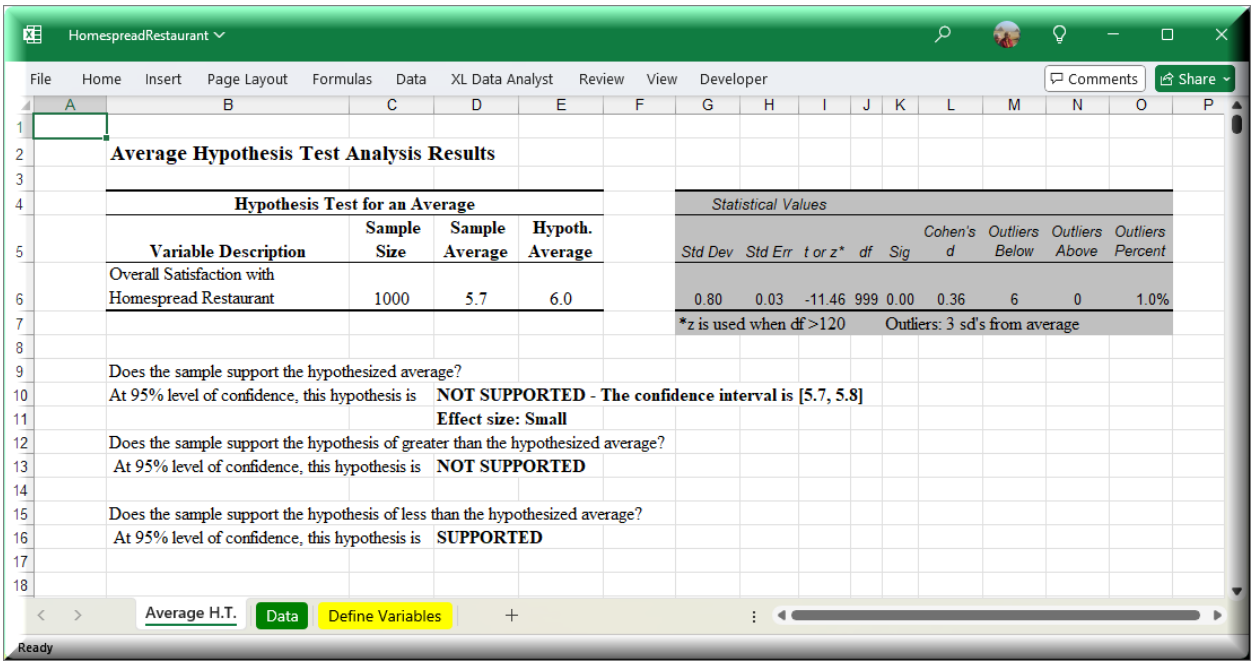
Options: There is one option available on the Hypothesis Test - Average window:

- Remove Outliers? – The user may opt to have any outliers removed from the analysis if found. The default is "No"

Discovery: The XL Data Analyst "Average H.T.#" output worksheet identifies the variable selected, its sample size, sample average, and the hypothesized population average for that variable. Beneath the table are found insights as to support (SUPPORTED) or nonsupport (NOT SUPPORTED) for the hypothesized average tested at the specified % level of statistical significance. In the case of nonsupport, the confidence interval (at specified % level of confidence) is reported.

Beneath the statement of support or nonsupport for the precise hypothesis, there is a statement for the "greater than" hypothesis support or nonsupport and a statement for the "less than" hypothesis support or nonsupport with respect to the hypothesized value.

To the right of the discovery table is a Statistical Values table for the average hypothesis test. Specifically, standard deviation, standard error, computed z (or t) value, degrees of freedom, and significance level are listed in this table. These values can be ignored by most users, whereas advanced users may wish to examine them. Outlier information is also contained in the Statistical Values table.

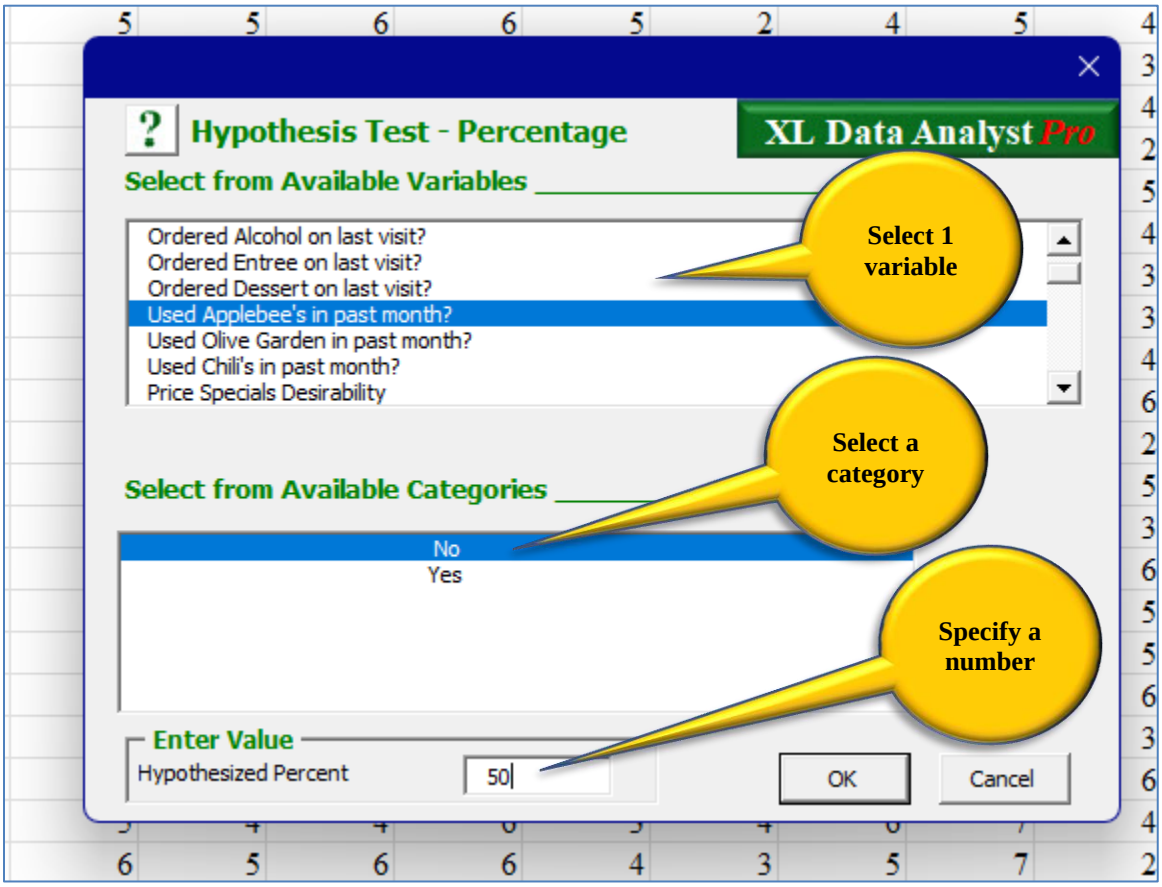


Average Hypothesis Test Discovery

Hypothesis Test for a Percent

With a hypothesis test, the analyst uses a random sample's findings to "support" or "not support" a belief (hypothesis) as to the population percent for some category within a categorical variable. With a percent hypothesis test, the sample percent, sample size, standard error of the percent, and z value are used. Again, XL Data Analyst Pro incorporates the (e.g.) 95% level of confidence as default, so the result is a statement that the analyst is 95% confident of finding or not support for the hypothesis. The user may run this analysis with any desired level of confidence by using the "Significance Level" feature.

? % **Procedure:** Use the Hypothesis Test-Percent icon sequence to open the Hypothesis Test – Percentage window. Select a categorical variable in the "Select from Available Variables" window then select one of the value labels in the "Select from Available Categories" window. Note that in the case of a metric variable with no value labels, nothing will appear in the "Select from Available Categories" window. Specify the Hypothesized Percent according to the hypothesis being tested.

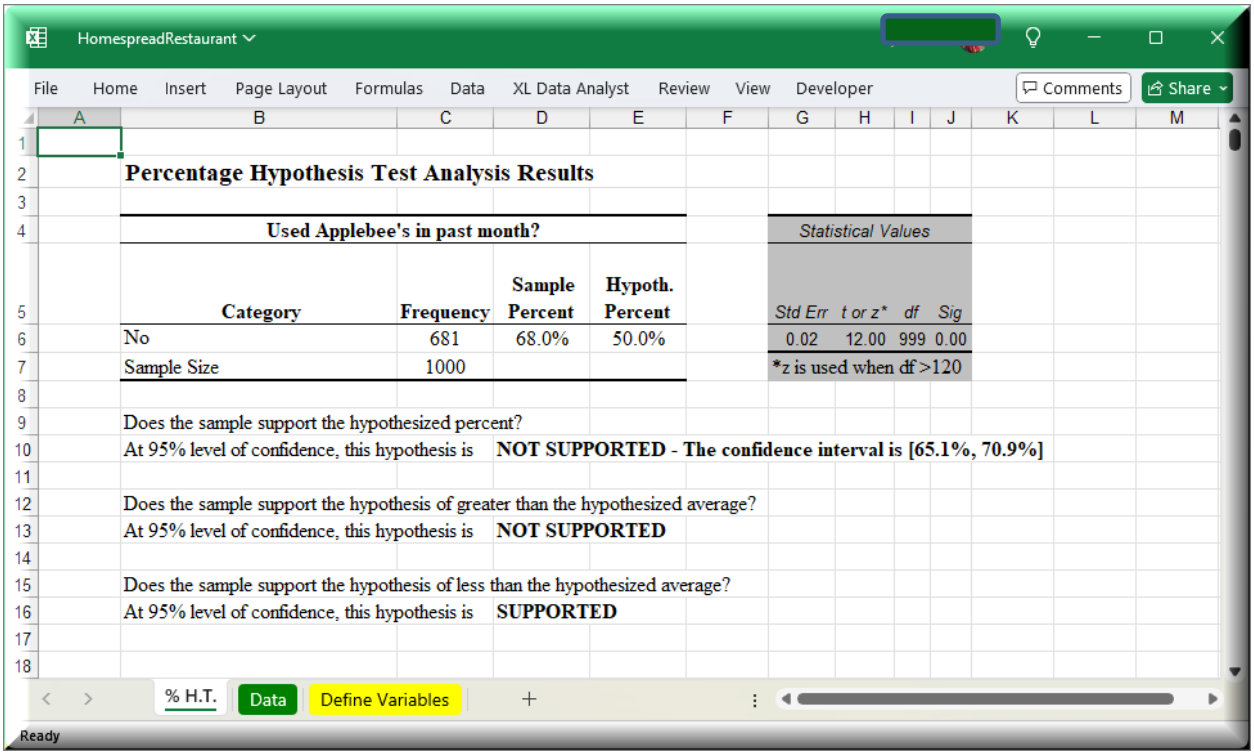


Percentage Hypothesis Test Window

Discovery: The XL Data Analyst Pro "% H.T.#" output worksheet table identifies the variable selected, the category being tested, its frequency, sample percent, and hypothesized percent for that category. Sample size is identified as well. Beneath the table are found insights as to support (SUPPORTED) or nonsupport (NOT SUPPORTED) for the hypothesized population percent tested at the specified % level of statistical significance. In the case of nonsupport, the confidence interval (at specified % level of confidence) is reported.

Beneath the statement of support or nonsupport for the precise hypothesis, there is a statement for the "greater than" hypothesis support or nonsupport and a statement for the "less than" hypothesis support or nonsupport with respect to the hypothesized value.

To the right of the discovery table is a Statistical Values table for the average hypothesis test. Specifically, standard deviation, standard error, computed z (or t) value, degrees of freedom, and significance level are listed in this table. Again, most users can ignore these values, while advanced users may wish to examine them.



Percentage Hypothesis Test Discovery

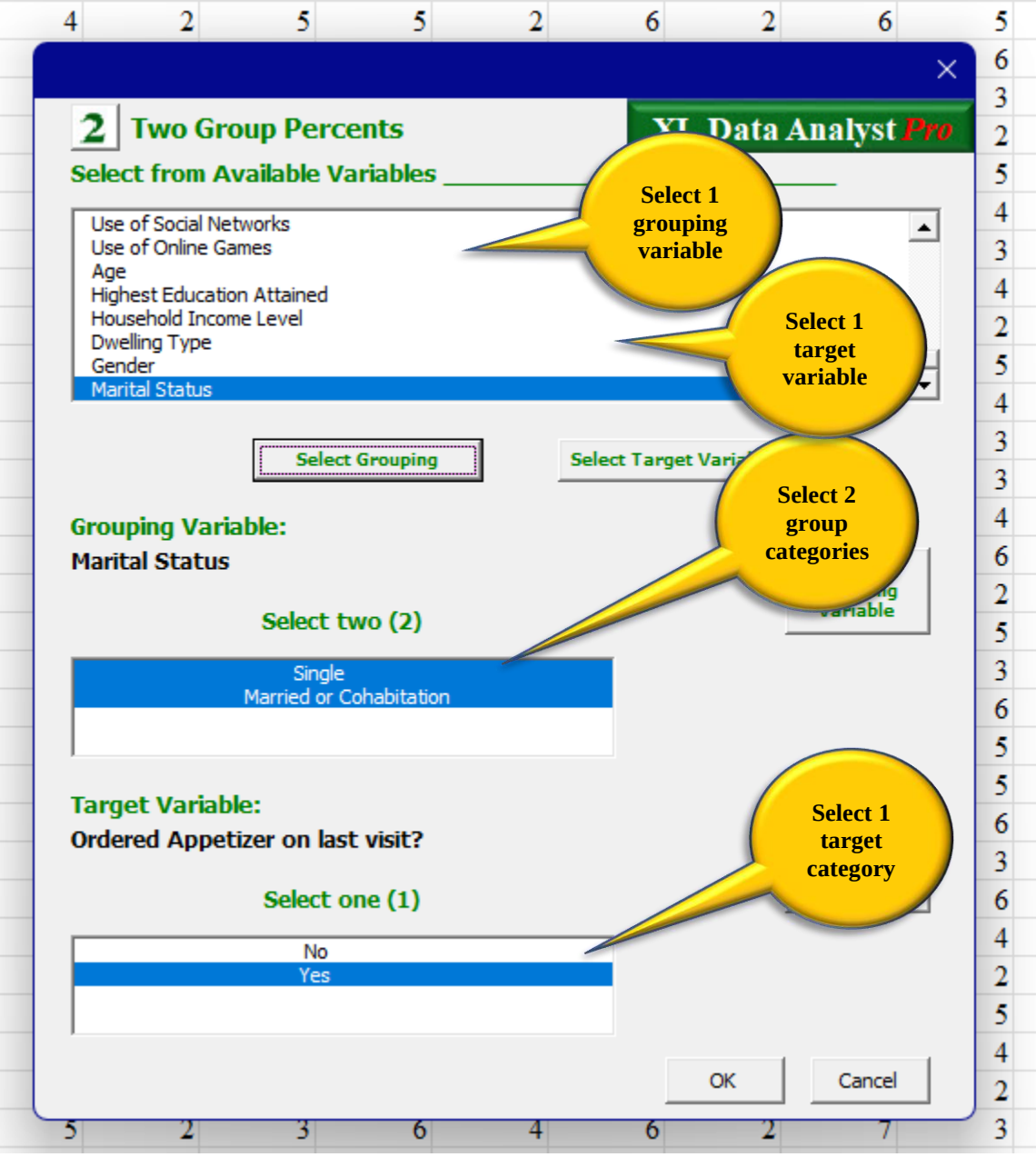
Differences Analyses

Differences analysis, found in the "Differences" XL Data Analyst Pro ribbon group, pertains to comparisons of groups. Typically, these groups are subpopulations within the larger total population such as males versus females. With the identified groups, the percentage or average value of each group for a particular target variable are assessed for statistically significant difference. If found, the analyst is (e.g.) 95% confident that the difference between the sample values (such as percent for males versus percent for females) exists in the population represented by the random sample. There are four differences test types possible: (1) difference in percents for each of two groups, (2) difference in averages for each of two groups, (3) differences in averages for three or more groups, and (4) difference in averages for two variables.

Differences between 2 Group Percents

Two groups in a categorical variable are identified in the population (such as males versus females or urban versus rural home dwelling locations) and percents are calculated for a categorical value response (such as percent indicating "yes" to owning Amazon Echo). Using sample percents, sample sizes, standard error of the percent difference, and (e.g.) 95% level of confidence z of (e.g.)1.96, a determination is made for support or no support for the hypothesis that the population difference between these two percents is 0. When the hypothesis is supported, the analyst concludes that no statistically significant difference exists between the two groups. On the other hand, with no support for the hypothesis, the analyst concludes, at (e.g.) 95% level of confidence that a statistically significant difference does exist between the two groups in the population. With XL Data Analyst Pro, the user can specify any desired level of confidence for this test.

2 % **Procedure:** Use 2 Groups-Percents to open up the Two Group Percents selection window. Highlight the categorical grouping variable in the "Select from Available Variables" window and click the "Select Grouping" button. With highlighting, select two group categories from those that appear in the "Select two (2)" window. Highlight the categorical target variable in the "Select from Available Variables" window and use the "Select Target" button to select it. Then identify with highlight the desired category from those listed in the Target Variable/Select one (1) window.

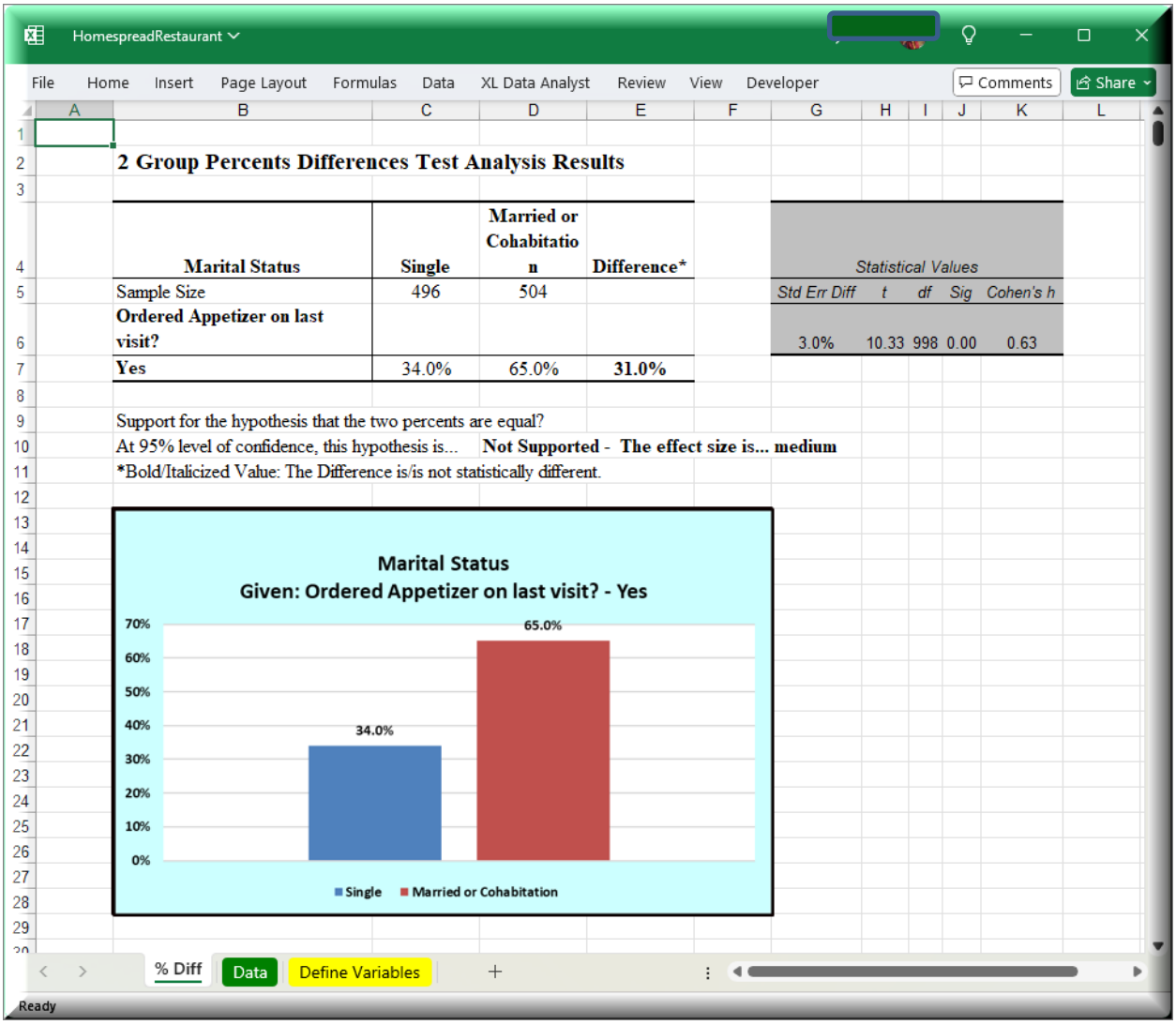


2 Group Percents Difference Window

Discovery – Findings & Data Visualization: XL Data Analyst Pro produces a "% Diff#" output worksheet with a table identifying the two groups in the grouping variable as columns and relates the category chosen in the target variable. The frequencies and the percents of each of the two groups are specified, and the arithmetic difference between the two percents is given. Beneath the table, the result of the specified % level of confidence hypothesis test identifies whether the hypothesis (null hypothesis that the two percents are equal) is supported or not supported. As a further interpretation aid, the percent difference value is bolded if it is statistically significant and italicized if not. If the hypothesis is not supported, meaning that the difference between the two percents is statistically significant, the effect size is reported as trivial, small, medium, or large based on Cohen's h guidelines.

A Statistical Values table is placed alongside this table, and it reports: standard error of the difference in percents, t or z, degrees of freedom, significance level, and Cohen's h.

If a statistically significant difference in percents occurs, XL Data Analyst Pro creates a visual comparison of the percents with a column graph.

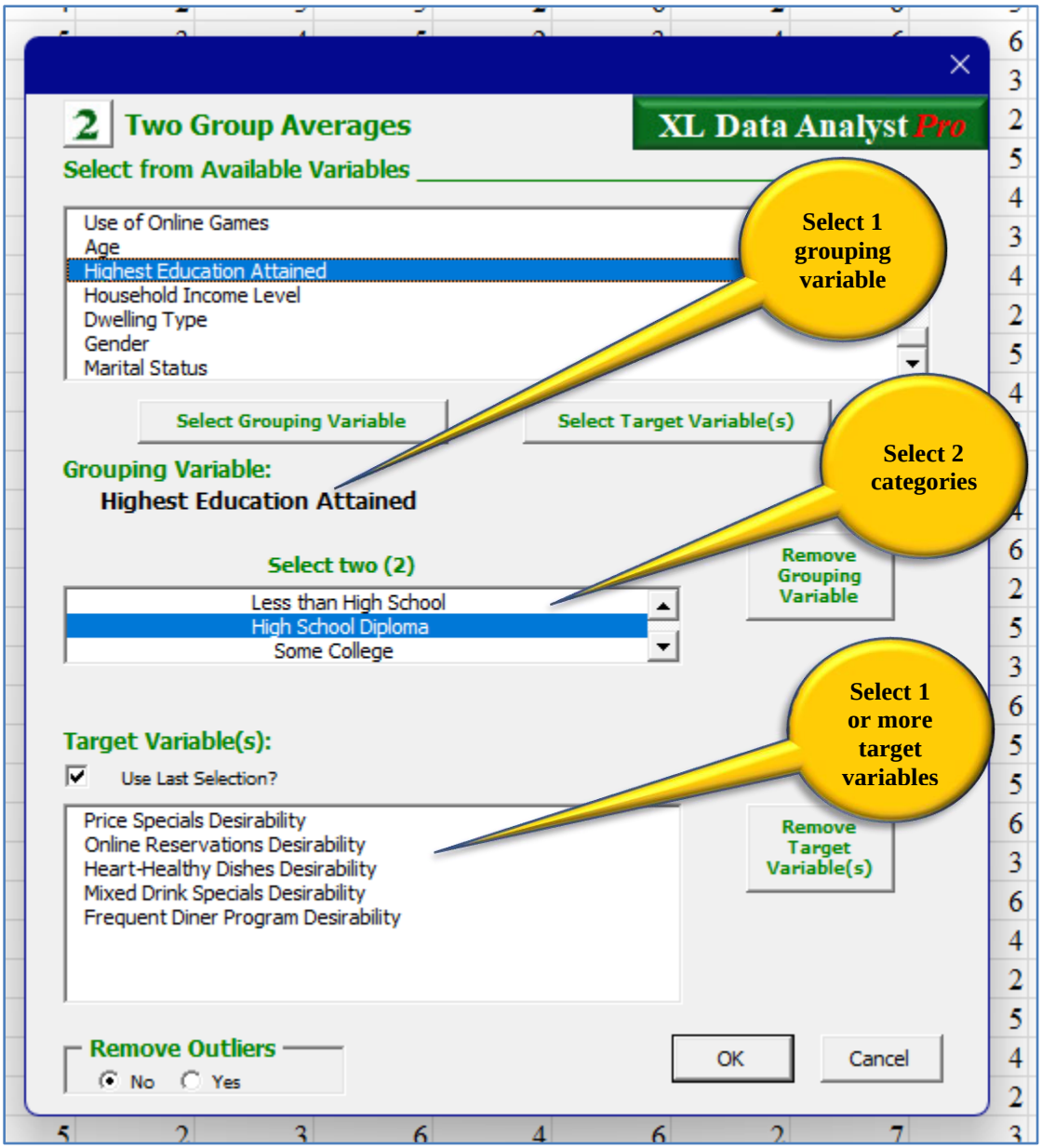


2 Group Percents Difference Discovery

Difference between 2 Group Averages

Two groups are identified in the population (such as males versus females or urban versus rural home dwelling locations) and group averages are calculated for some metric variable (such as average amount of time on Facebook daily). Using sample means, sample sizes, standard deviation, standard error of the average difference, and (e.g.) 95% level of confidence z of (e.g.) 1.96, a determination is made for support or no support for the (null) hypothesis that the population difference between these two averages is 0. When the null hypothesis is supported, the analyst concludes that no statistically significant difference exists between the two groups. On the other hand, with no support for the hypothesis, the analyst concludes, at (e.g.) 95% level of confidence, that a statistically significant difference does exist between the two groups in the population. With XL Data Analyst Pro, the user can specify any desired level of confidence for this test.

Procedure: Use the 2 Groups – Averages icon sequence to open the Two Group Averages selection window. Highlight the categorical grouping variable in the "Select from Available Variables" window and click the "Select Grouping" button. With highlighting, select two group categories from those that appear in the "Select two (2)" window. Highlight the metric target variable(s) in the "Target Variable(s)" window. Note that multiple target variables can be selected.



2 Group Averages Difference Window

Options: There are two options available on the Two Group Averages Difference window:

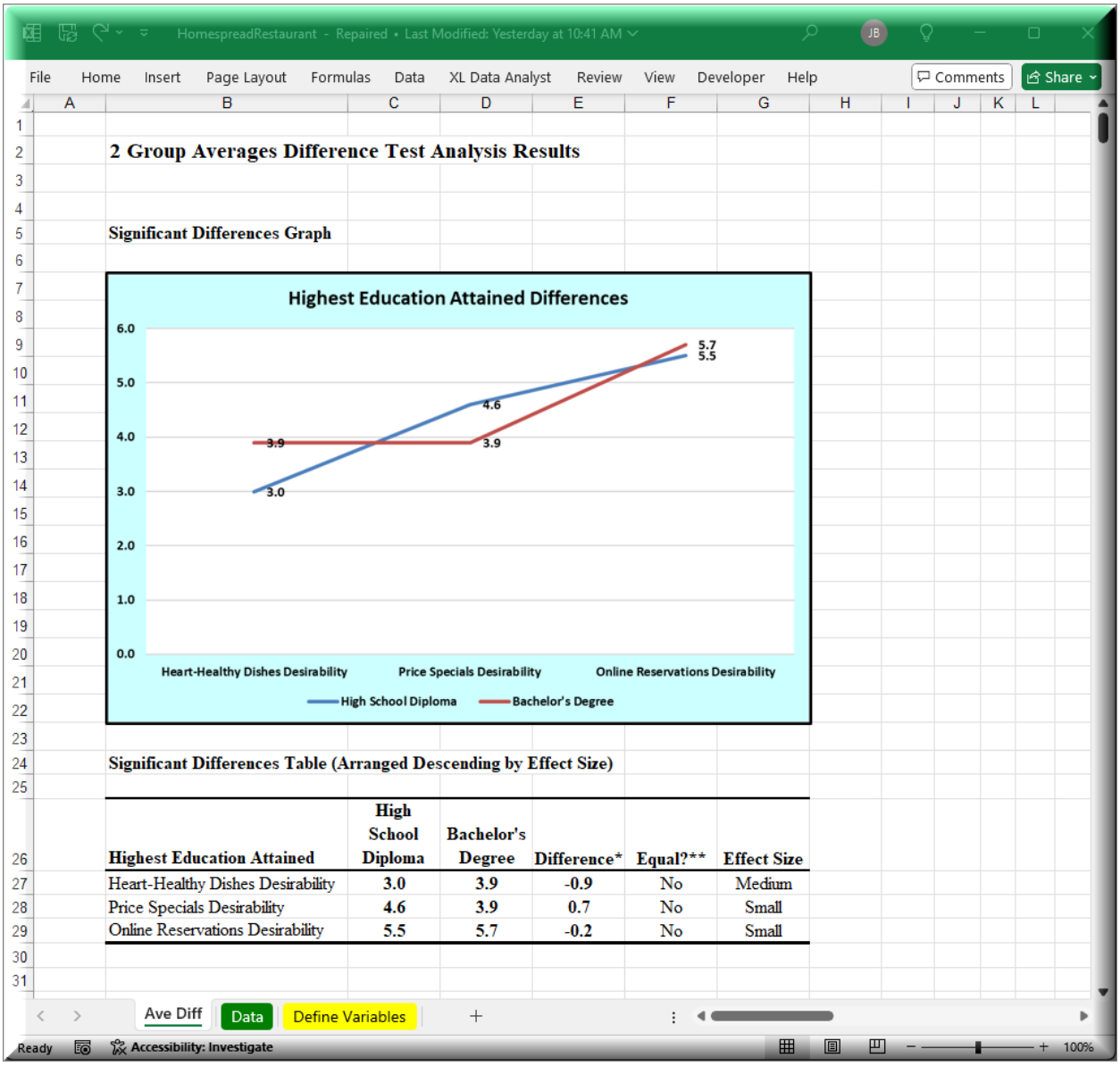
- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.
- Remove Outliers? – The user may opt to have any outliers removed from the analysis if found. The default is "No"

Discovery – Data Visualization: The XL Data Analyst produces an "Ave Diff#" output worksheet with three possible parts: visualization, significant findings, and findings. If XL Data Analyst Pro determines that two or more average differences are statistically significant, it creates a visualization in the form of a comparative line graph; whereas, if only one pair of averages is significantly different, it creates a bar graph. The graph identifies each group and plots the group averages for each variable where a statistically significant difference occurs.

Discovery – Significant Findings: Immediately following the graph is an identified table of only variables where statistically significant average differences are found. The table lists the averages and sample sizes of each of the two groups as well as the arithmetic difference between the averages. Because all variables in this table have statistically significant differences, a "No" is specified in the "Equal?**" column. The last column reports the effect size for the significantly different averages for that variable. Effect size is computed via Cohen's d, and the possible size levels are: "None," "Very small," "Small," "Medium," "Large,", "Very large," and "Huge." The variables in this table are arranged from highest effect size to lowest effect size.

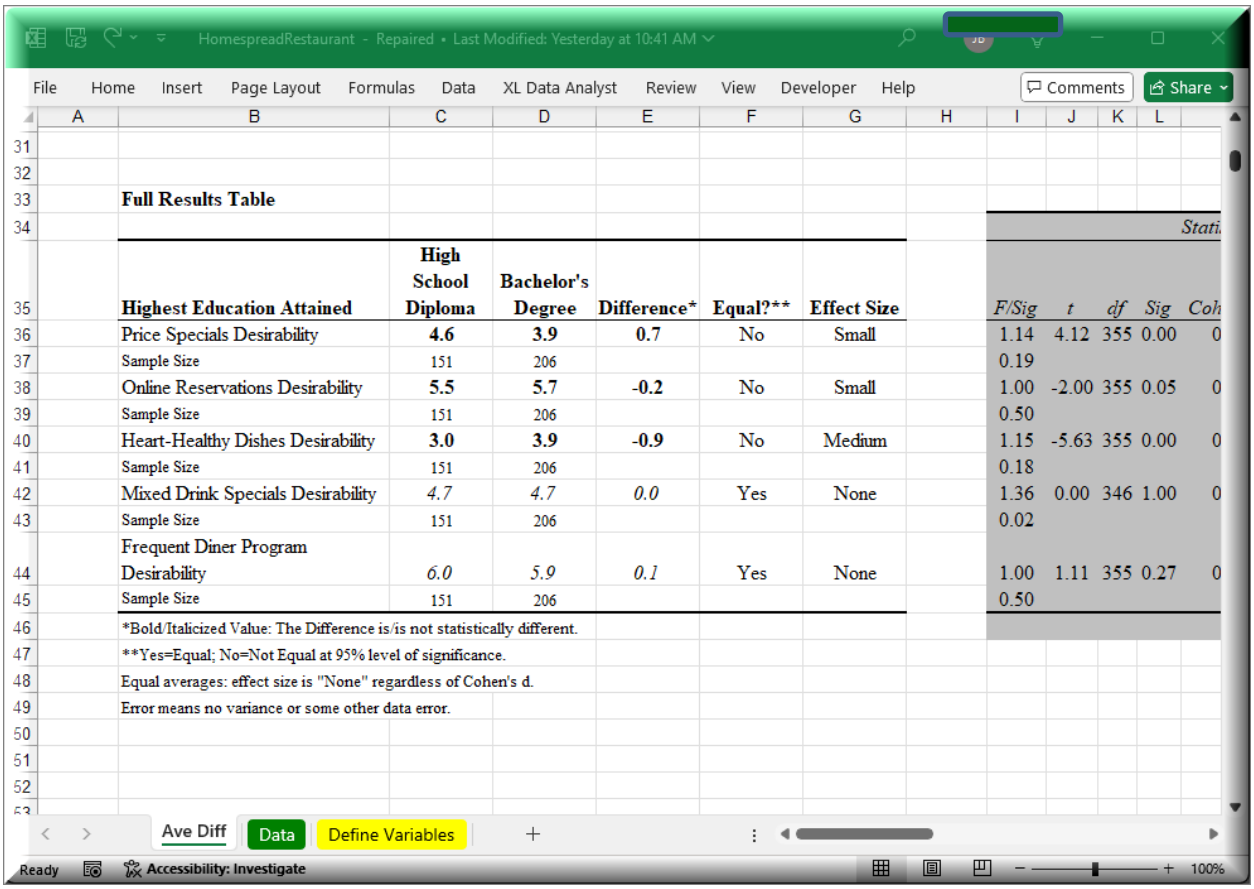
Discovery – Findings: Xl Data Analyst Pro always produces a table identifying the two groups in the grouping variable as columns and identifies chosen target variable(s). The averages and sample sizes of each of the two groups are specified, and the arithmetic difference between them listed. Because more than one target variable may be in the table, and multiple average differences hypothesis tests involved, the table has a column identified as "Equal?**" that specifies with "Yes" or "No" the result of the hypothesis test for each pair of averages. "Yes" signifies no support for the null hypothesis or the existence of statistically significant difference between the two averages involved (i.e., the two averages are statistically equal), while "No" indicates statistically significant difference exits for that pair of averages (i.e., the two averages are statistically different). The last column in the table reports effect sizes of significantly different differences as described above.

A separate greyed Statistical Values table, of possible interest to advanced users, is placed alongside this table, and in it are found Statistical Values for: F value and its significance level, computed t value, degrees of freedom, and significance level corresponding to each hypothesis test. The procedure uses Levine's test for the equality of the variances of the two groups as a means of identifying the proper equation to use for the computation of the t value. Values for Cohen's d are reported in this table as well as the results of outliers analysis.



Highest Education Attained	High School Diploma	Bachelor's Degree	Difference*	Equal?**	Effect Size
Heart-Healthy Dishes Desirability	3.0	3.9	-0.9	No	Medium
Price Specials Desirability	4.6	3.9	0.7	No	Small
Online Reservations Desirability	5.5	5.7	-0.2	No	Small

2 Group Averages Data Visualization and Significant Differences



Full Results Table

Highest Education Attained	High School Diploma	Bachelor's Degree	Difference*	Equal?**	Effect Size	F/Sig	t	df	Sig	Coh
Price Specials Desirability	**4.6**	**3.9**	**0.7**	No	Small	1.14	4.12	355	0.00	0
Sample Size	151	206				0.19				
Online Reservations Desirability	**5.5**	**5.7**	**-0.2**	No	Small	1.00	-2.00	355	0.05	0
Sample Size	151	206				0.50				
Heart-Healthy Dishes Desirability	**3.0**	**3.9**	**-0.9**	No	Medium	1.15	-5.63	355	0.00	0
Sample Size	151	206				0.18				
Mixed Drink Specials Desirability	*4.7*	*4.7*	*0.0*	Yes	None	1.36	0.00	346	1.00	0
Sample Size	151	206				0.02				
Frequent Diner Program Desirability	*6.0*	*5.9*	*0.1*	Yes	None	1.00	1.11	355	0.27	0
Sample Size	151	206				0.50				

*Bold/Italicized Value: The Difference is/is not statistically different.

**Yes=Equal; No=Not Equal at 95% level of significance.

Equal averages: effect size is "None" regardless of Cohen's d.

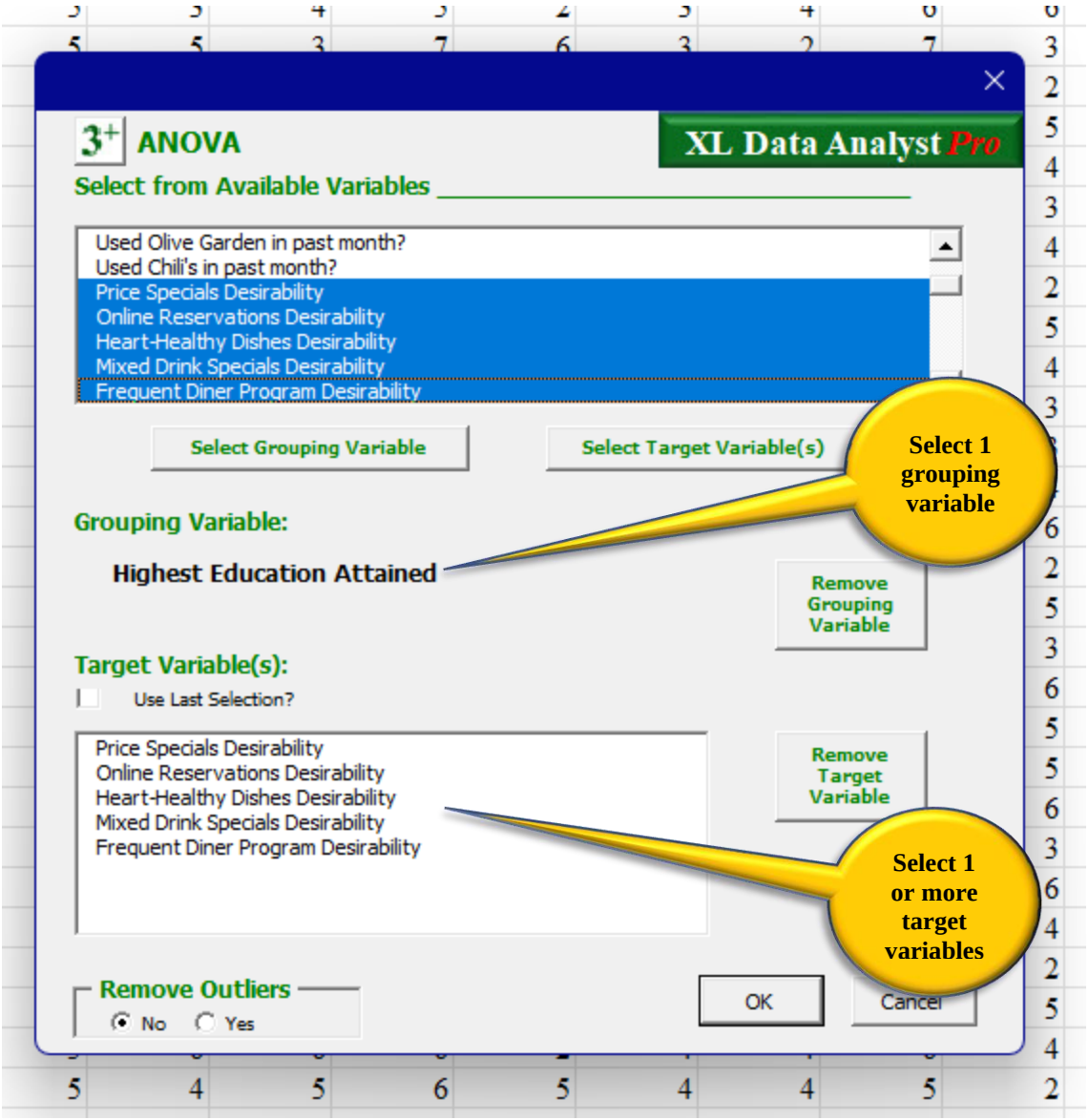
Error means no variance or some other data error.

2 Group Averages Discovery

Differences among 3+ Groups (ANOVA)

Three or more groups are identified in the population and the average for each group is calculated for some metric variable. One-way analysis of variance (ANOVA) procedures are applied, resulting in an assessment as to whether any two groups' averages difference is significantly different (no support for the hypothesis that the difference is zero). If a significant difference is found, one of several possible post hoc procedures may be used to identify precisely what group average(s) differ from what other group average(s). XL Data Analyst Pro uses Scheffe's Test due to its somewhat a conservative nature and simultaneous confidence intervals (somewhat less conservative) for the data visualization graph and ease of interpretation of its presentation of the findings.

3+ **Procedure:** Use the 3+ Group Averages icon to open the ANOVA variable selection window. Highlight a categorical variable in the "Select from Available Variables" window as the grouping variable and use the "Select Grouping" button to identify it as the Grouping Variable. There is no need to identify the groups as XL Data Analyst Pro will find all eligible groups in the selected grouping variable. Use highlighting in the "Select from Available Variables" window and the "Add Target Variable(s)" button to select one or more metric target variables into the "Target Variable(s)" window.



3+ Group Averages (ANOVA) Window

Options: There are two options available on the ANOVA variables selection window:

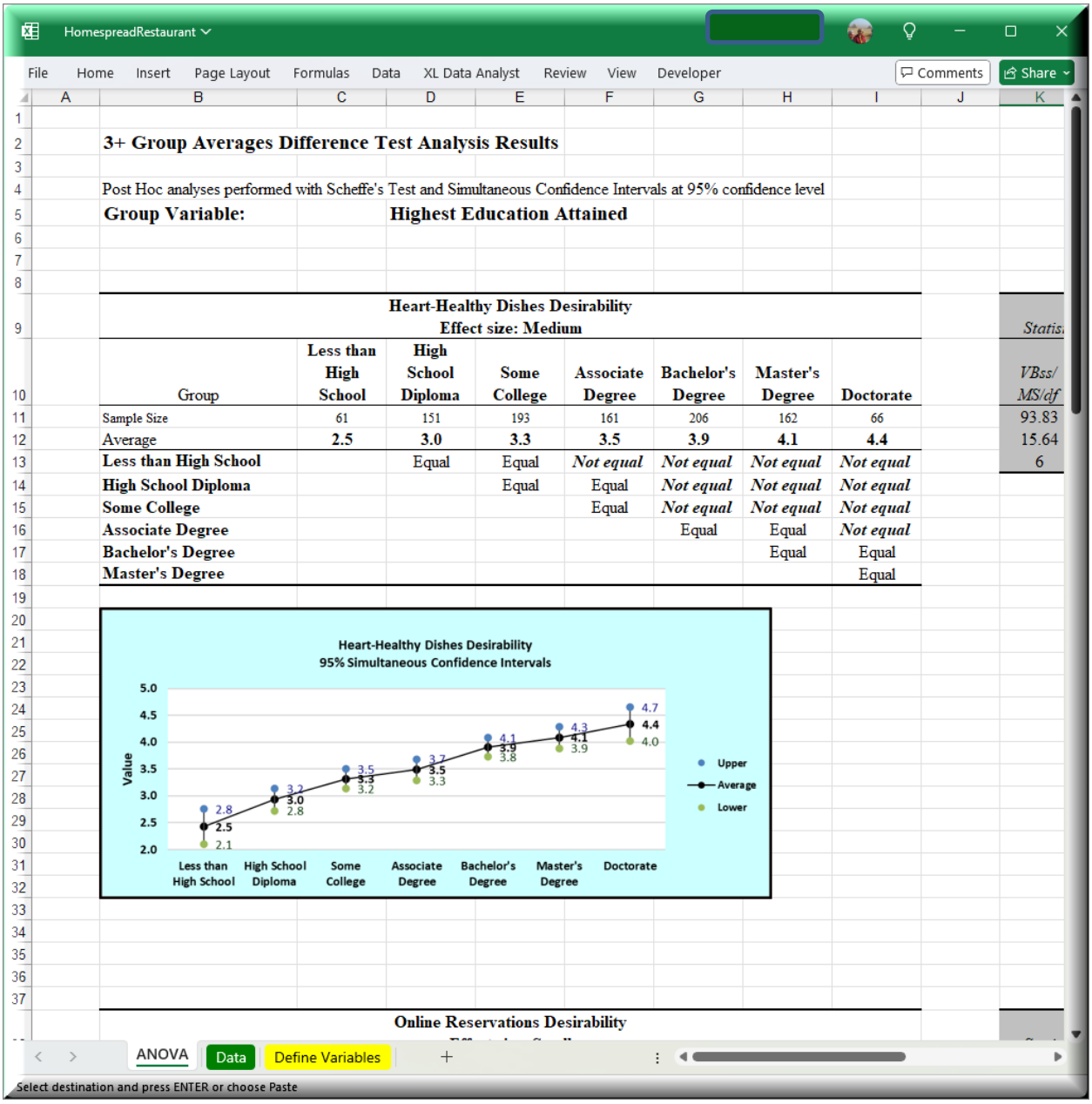
- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.
- Remove Outliers? – The user may opt to have any outliers removed from the analysis if found. The default is "No"

Discovery – Findings: An "ANOVA#" output worksheet is produced, and each target variable's analysis is contained in a separate table. Each table identifies the groups in the grouping variable as rows and columns and lists the sample size and average for the target variable for each group in the grouping variable. The groups are arranged in the table by size of their averages, low to high, from left to right and from top to bottom. If warranted by the overall ANOVA F test, the table matrix cells reveal the result of each pair of averages difference test, using Scheffe's post hoc test identified with "Equal" or "Not Equal." In addition, the effect size based on eta squared is reported near the top of the table. Possible effect size descriptors are: "Less than small," "Small," "Medium," and "Large."

If the overall ANOVA F test results in a significance level of greater than the specified level, the table lists only sample sizes and averages for the groups with a note of no significant differences at the specified level of confidence.

For each ANOVA, the worksheet has a greyed Statistical Values table is placed alongside each target variable table (only partially shown here) listing values for the overall ANOVA test values – sums of squares, mean square, F test value, and significance level. It also holds the eta squared value and outliers analysis results.

Discovery – Data Visualization: If a statistically significant finding is discovered, XL Data Analyst Pro creates a line graph of the group averages and (e.g.) 95% simultaneous confidence intervals as a visual aid. The graph is placed immediately below the findings table.

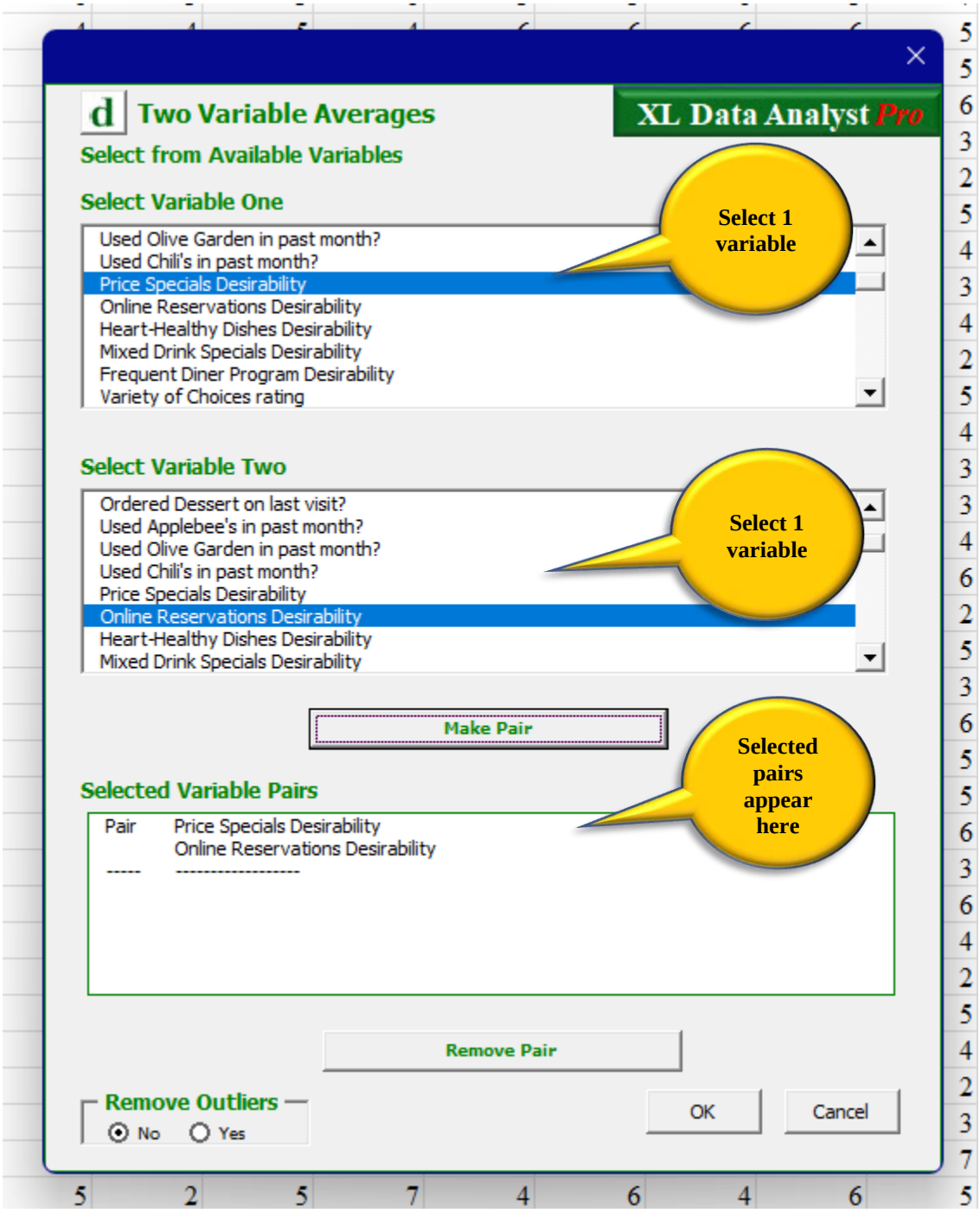


3+ Group Averages Differences (ANOVA) Discovery and Data Visualization

Difference between 2 Variable Averages

As opposed to the averages of groups, the averages of 2 variables are assessed for statistically significant difference. Obviously, the variables must be of the same scale (such as both in dollars, or amount of times, or responses to a 5-point scale of "strongly disagree" to "strongly agree") for the differences to make sense. The sample averages, sample sizes, standard deviations, standard errors, and (e.g.) 95% level of confidence are applied in the analysis to determine whether there is support or no support for the hypothesis that the difference between the two averages is zero in the population. With no support, the analyst can attest that the difference found in the sample exists in the population.

d **Procedure:** Use the Paired Differences icon to open the Two Variable Averages selection window. Select a variable from the "Select Variable One" window pane, and another variable from the "Select Variable Two" pane. The "Make Pair" button places the selected pair in the "Selected Variable Pairs" window. Multiple pairs can be selected.



2 Variable Averages Variables Selection Window

Options: There is one option available on the ANOVA variables selection window:

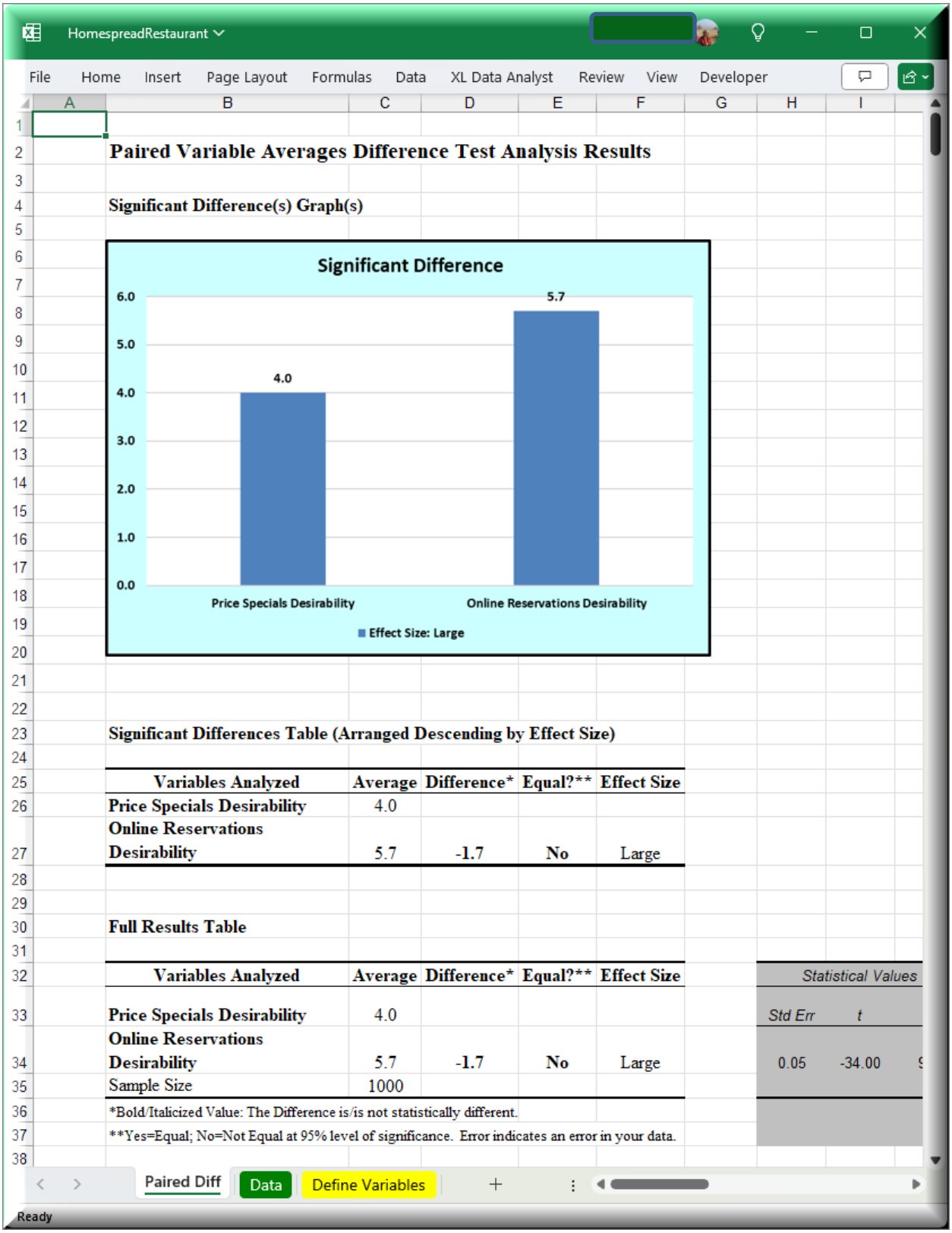
- Remove Outliers? – The user may opt to have any outliers removed from the analysis if found. The default is "No"

Discovery – Data Visualization: The XL Data Analyst produces a "Paired Diff#" output worksheet, and if statistically significant findings are discovered, it produces a comparative bar chart for each set of significantly different averages.

Discovery – Significant Findings: Immediately following the graph is an identified table of only pairs of variables where statistically significant differences in their averages are found. The table lists the averages of each of the two groups as well as the arithmetic difference between the averages. Because all variables in this table have statistically significant differences, a "No" is specified in the "Equal?**" column. The last column reports the effect size for the significantly different averages for that variable. Effect size is computed via Cohen's d, and the possible size levels are: "None," "Very small," "Small," "Medium," "Large,", "Very large," and "Huge." The variables in this table are arranged from highest effect size to lowest effect size.

Discovery – Findings: The full results table reports the average of each variable in each selected pair and the sample size. The arithmetic difference is listed and in the "Equal?*" column, it specifies "Yes" or "No" indicating if the hypothesis test of no difference between the variables' averages (null hypothesis) is supported or not supported, respectively. The last column reports the effect size of the difference between the two averages.

The accompanying greyed Statistical Variables table (only partially shown here) provided for advanced users identifies: standard error of the difference, computed t value, degrees of freedom, and significance level for each pair. It reports Cohen's d values and the findings of outlier analysis.



2 Variable Averages Difference Data Visualization and Discovery

Relationship Analyses

Relationship analyses seek to assess the degree to which a systematic pattern or association exists between two or more variables. Depending on what analysis is performed, the analyst can assess relationship stability, strength, and/or directionality. Stability refers to whether the relationship exists in the population; strength pertains to the magnitude of the relationship as for instance a high correlation, while directionality refers to the positive or negative nature of the association. XL Data Analyst Pro performs three relationship analyses. Crosstabulation is applicable when analyzing two categorical variables; correlation is used when investigating the relationship between two metric variables, and multiple regression analysis can be applied to determine in what ways significant independent metric (mostly) variables are related to a single metric dependent variable.

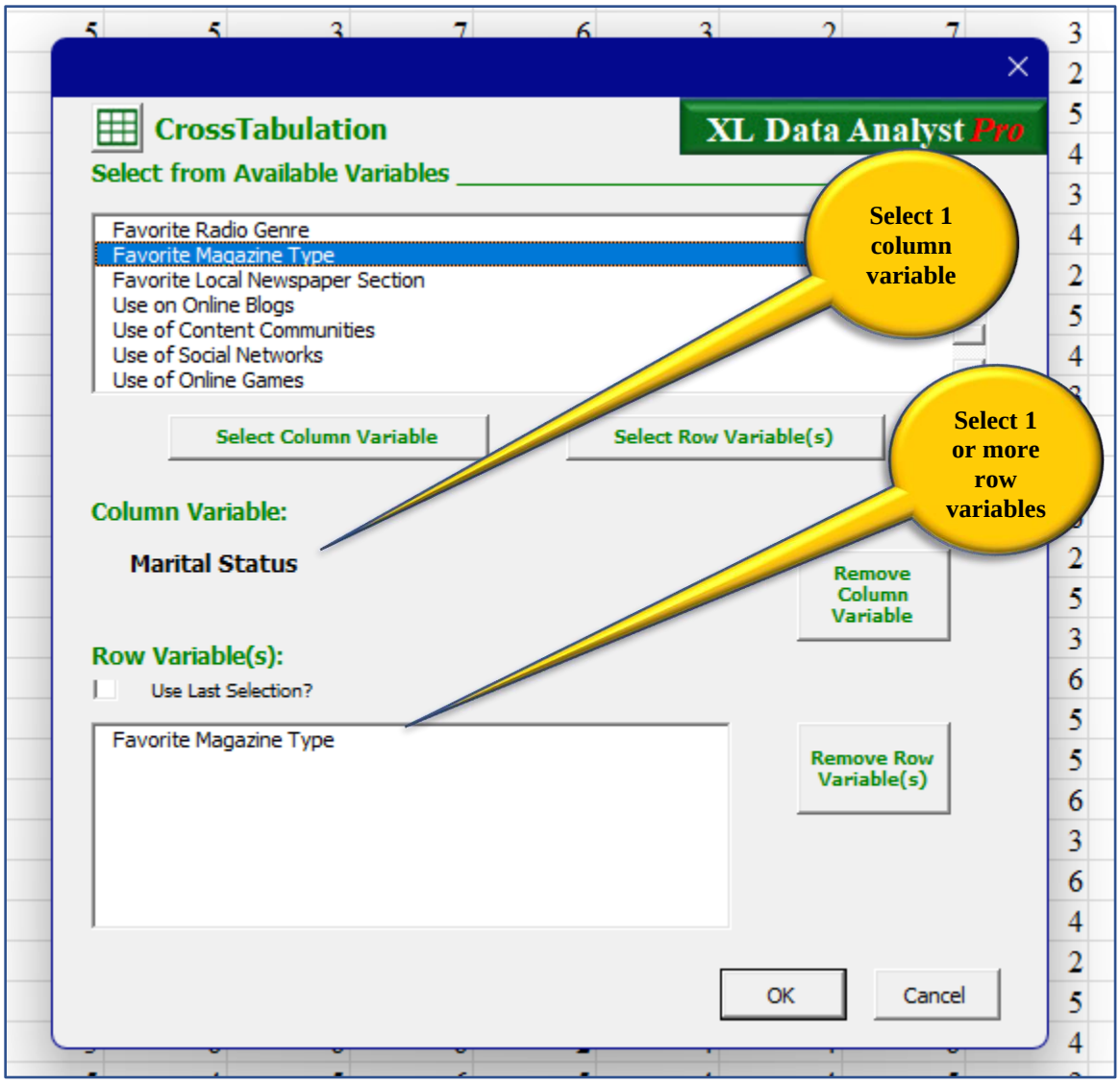
Crosstabulation Analysis

Crosstabulation analysis investigates the association between two categorical variables. The variables are arranged in a cross-classification table, meaning that one variable's categories are rows while the other variable's categories are the columns (does not matter which is which). For example, gender (categories: male versus female) and membership in Amazon Prime (categories: yes versus no) could be used, but multichotomous categorical variables are permitted.

The table forms a matrix, and the cells contain the number (frequency) of each pairing such as males who are members of Amazon Prime. Row totals, column totals, and a grand total are used. These frequencies, called "observed frequencies," are compared to frequencies that are "expected" if the data were to be distributed equitably across the matrix. For example, with a sample of 100 and 50/50 distribution of males and females plus membership/non membership, the individual cell frequencies would be expected to be 25.

Using the Chi Square formula, a Chi Square value is computed that expresses the deviation of the observed frequencies from the expected frequencies. If the computed Chi Square value exceeds a table value at the (e.g.) 95% level of confidence, the analyst concludes that there is a statistically significant association (relationship) between the two categorical variables. Due to the categorical nature of the variables involved, it is necessary to inspect percentage distributions, often depicted in graphs for ease of interpretation, in order to articulate the relationship(s) that exist in the population.

 **Procedure:** Clicking the Crosstabulation icon will open the CrossTab variables selection window. Use the "Select Column Variable" to choose a categorical variable from the "Select from Available Variables" window. One or more categorical Row Variables can be selected similarly using the "Add Row Variable(s)" button to specify the variable in the "Row Variable(s)" selection window. No categories or value labels need be specified for any variables as XL Data Analyst Pro determines the eligible categories as part of its analysis.

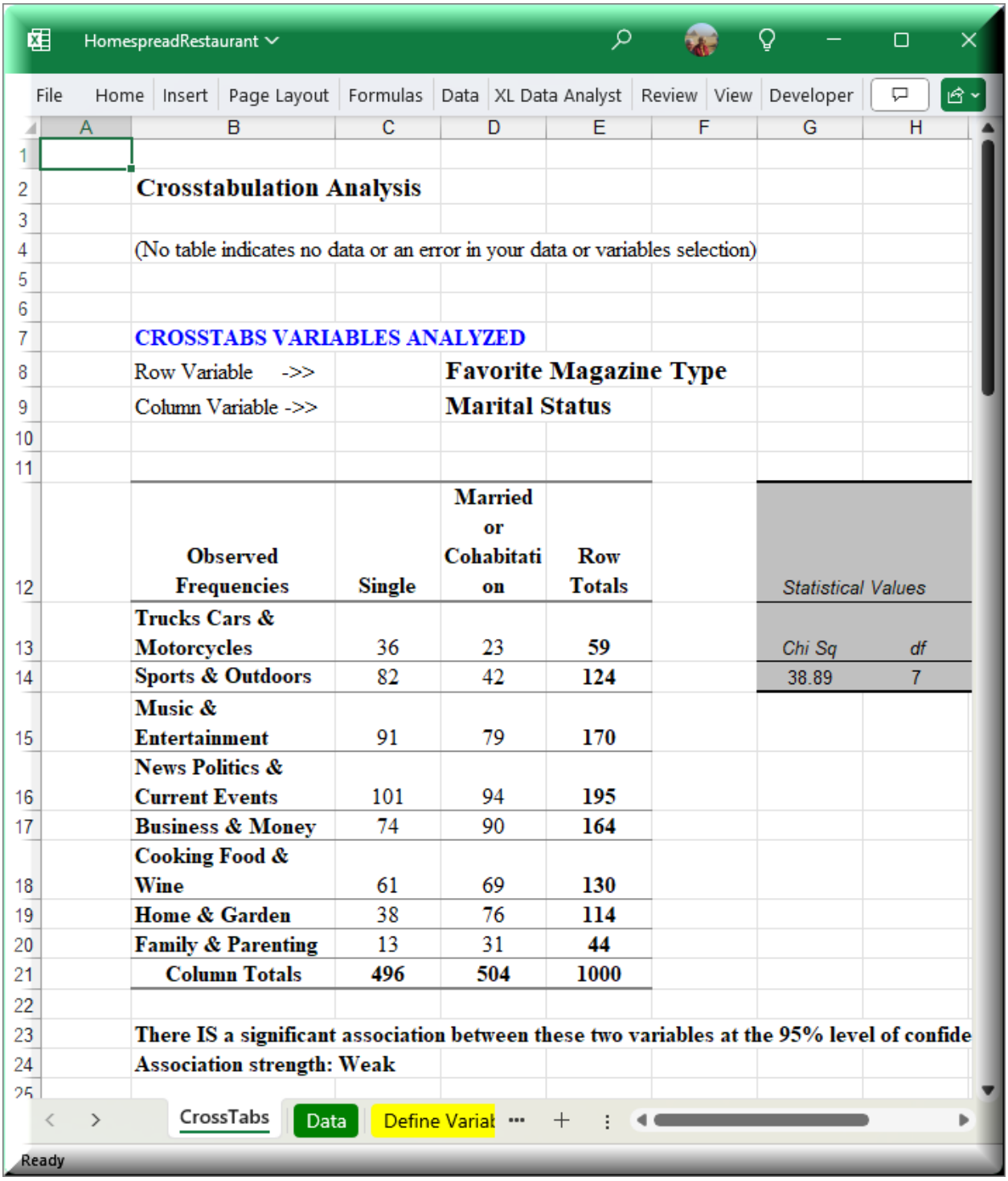


Crosstabulation Variables Selection Window

Options: There is one option available on the CrossTab variables selection window:

- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.

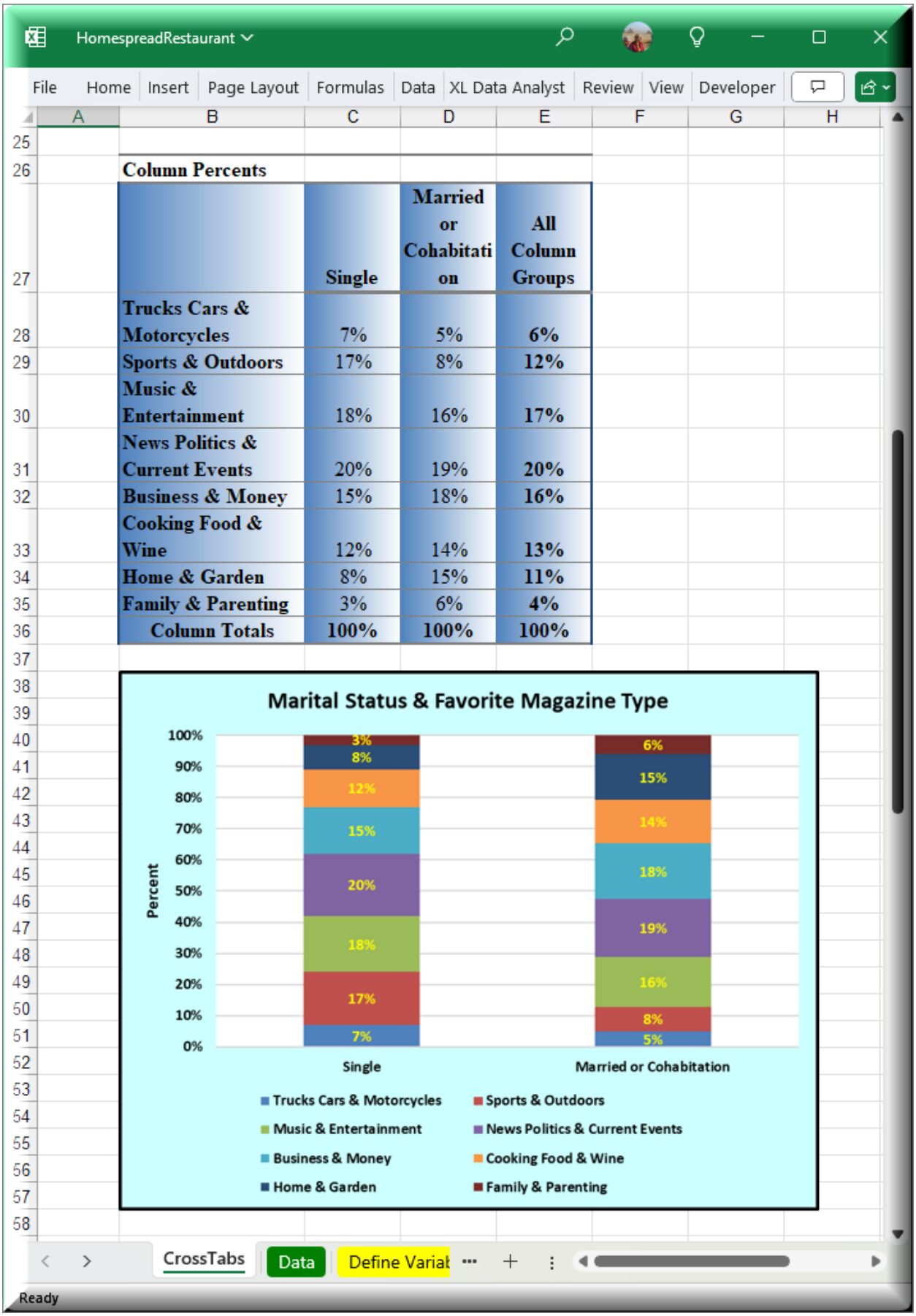
Discovery: On a "CrossTabs#" output worksheet, the XL Data Analyst provides a crosstabulation table of observed frequencies for each Column-Row variable pair. The Column Variable's categories are columns while the Row Variable's categories are rows in the crosstabulation table. Row and column totals as well as the grand total are provided in the table. Chi Square values – computed Chi Square, degrees of freedom, and significance - are provided alongside in a Statistical Values table for advanced users. Below the Observed Frequencies, the XL Data Analyst indicates the result of the Chi Square test. If the null hypothesis of no association is not supported, the specification is "There IS a significant association between these two variables (specified level of confidence)." Also, when a statistically significant association is discovered, XL Data Analyst Pro specifies its strength based on Cramer's V and notes it as "Negligible," "Weak," "Moderate," "Relatively strong," or "Strong." If no support is found for a significant relationship, the statement uses "IS NOT".



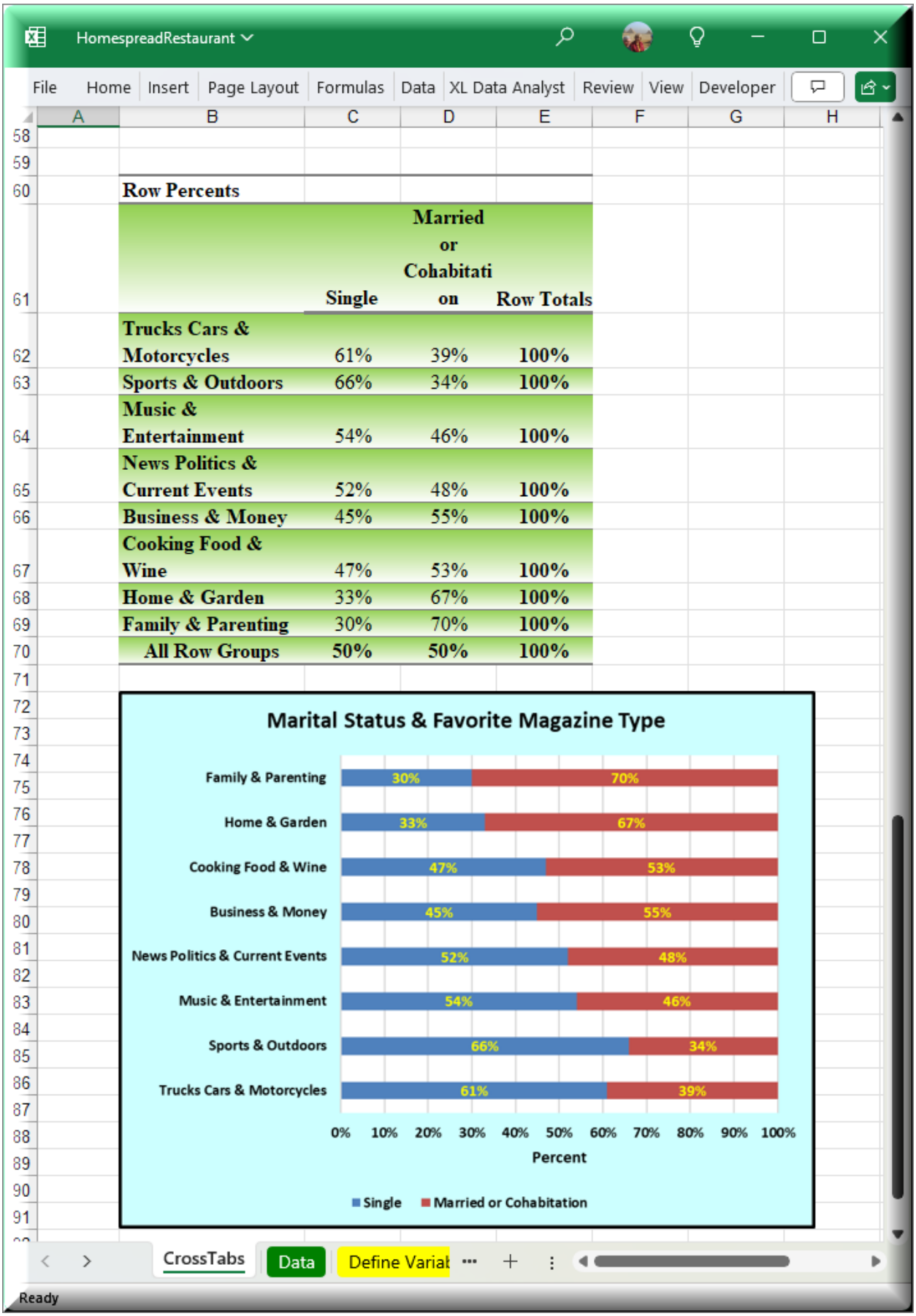
Crosstabulation Discovery

When XL Data Analyst Pro finds a significant association, it creates two additional tables below the Observed Frequencies table. One contains Column Percents (blue background color), while the other contains Row Percents (green background color). As an interpretation aid, the color shading in each table suggests the direction of the percents with respect to summing to 100%.

Discovery - Data Visualization: For the Column Percents and the Row Percents table, XL Data Analyst Pro creates comparative bar graphs. These graphs appear directly beneath each respective table.



Crosstabulation Discovery: Column Percents Table and Graph

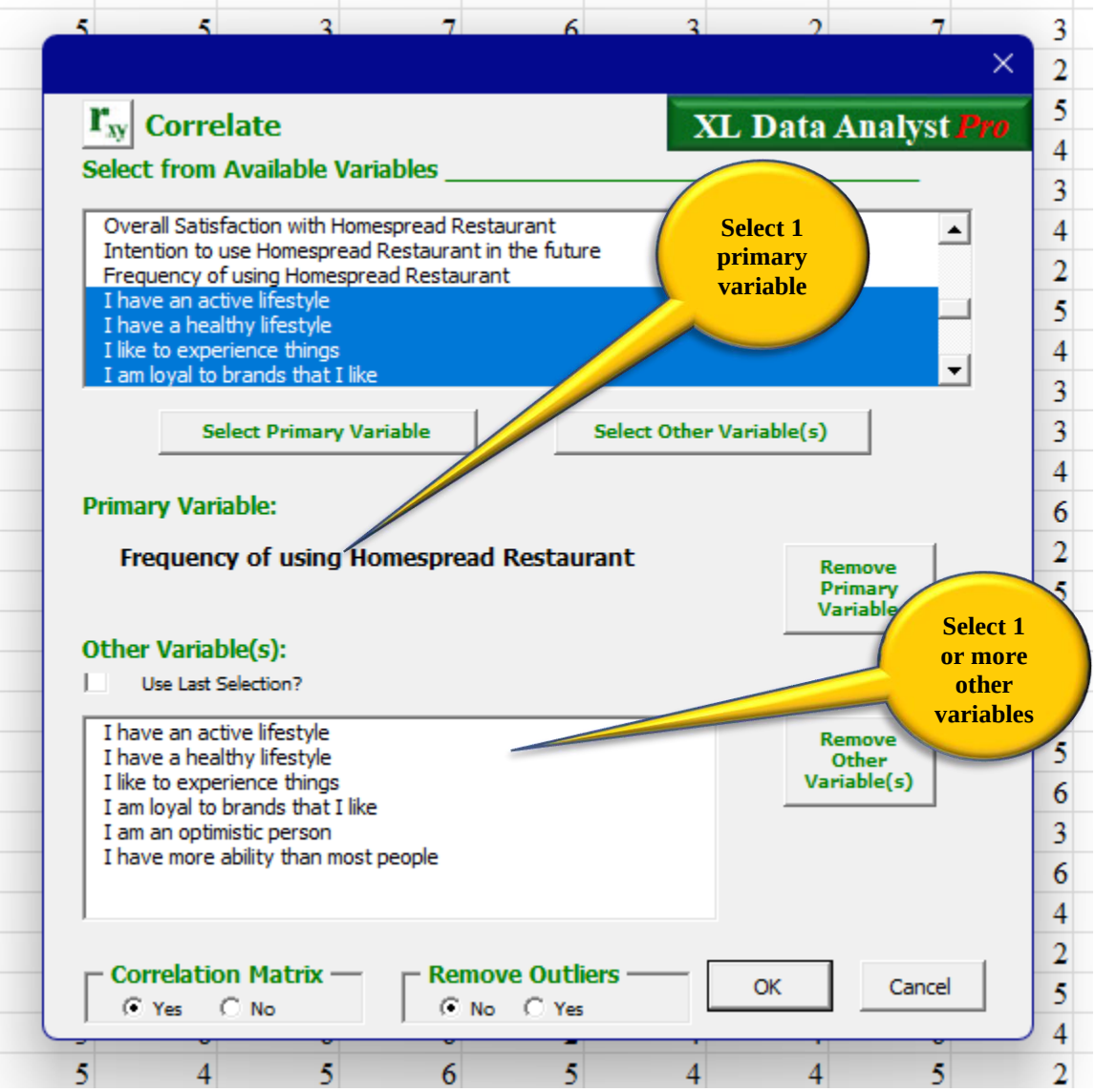


Crosstabulation Discovery: Column Percents Table and Graph

Correlation Analysis

Correlation is used to assess the association between two metric variables. Correlated metric variables "covary," meaning in the case of a positive correlation, as one variable increases, so does the other one. Whereas in the case of a negative correlation, as one variable increases, the other variable decreases in magnitude. A correlation coefficient ranges from -1 to +1 with 0 indicating no association whatsoever between the variables. The further away the correlation is from 0, either positive or negative, the stronger the correlation. Rules of thumb (refer to "Discovery" below) are sometimes used to label correlation strength. However, a computed correlation should be tested for statistical significance meaning that the population correlation is not zero (null hypothesis not supported). Sample size affects this test substantially. In other words, nonsignificant correlations are zero regardless of their magnitudes.

r_{xy} **Procedure:** Use the Correlation icon to open the Correlate variables selection window. A single, primary metric variable is selected with use of highlighting in the "Select from Available Variable" window and the "Select Primary Variable" button, while any number of metric Other Variable(s) can be selected with highlighting and the "Add Other Variable(s)" button.



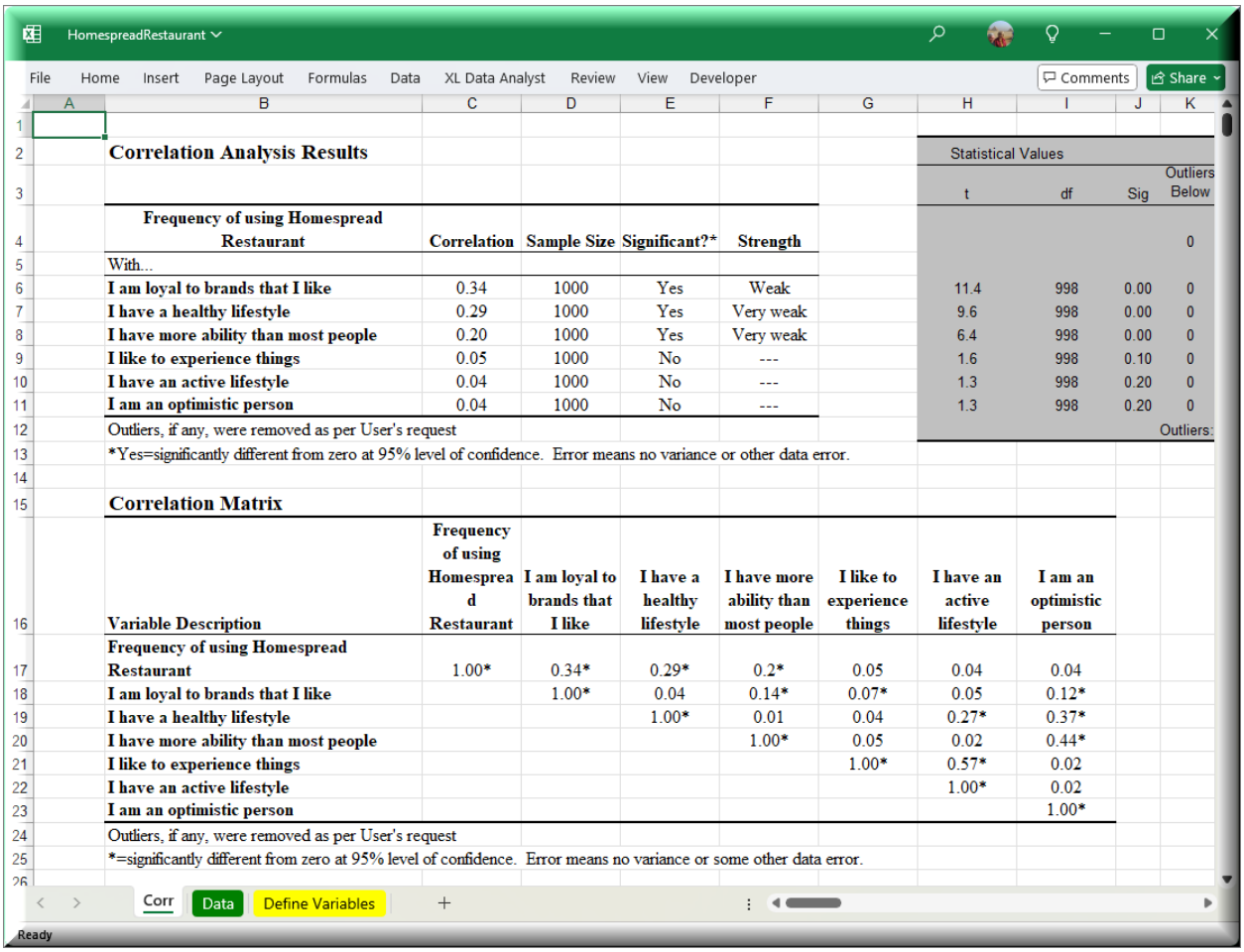
Correlate Variables Selection Window

Options: There are three options available on the Correlate variables selection window:

- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.
- Remove Outliers? – The user may opt to have any outliers removed from the analysis if found. The default is "No"
- Correlation Matrix – the correlation matrix for all variables will be produced along with a correlogram. The default is "Yes."

Discovery: With a "Corr#" output worksheet, XL Data Analyst Pro creates a table identifying the primary and other variables and indicating the Pearson Product Moment correlation, sample size, significance test result, and strength of the correlation for each pair. The "Other Variables" are listed in the table in descending order by size of their correlations, regardless of statistical significance, with the Primary Variable, positive to negative. The "Significant*" column uses "Yes" to indicate that the correlation is significantly different from zero in the population (null hypothesis of no correlation), and "No" if the hypothesis is supported. For those correlations found to be significant at the specified % level of confidence, correlation strength is indicated by a descriptive label according to rules of thumb based on the correlation coefficient's absolute value: Strong (above .81), Moderate (.61 to .8), Weak (.41 to .6), Very Weak (.21 to .4), or None (below .21).

If opted for, a Correlation Matrix is found below the correlations table. Only values in the upper diagonal of the matrix are reported as a correlation matrix is symmetric. Statistical significance at the specified % level of confidence of each correlation coefficient is indicated with an asterisk.



Correlation Analysis Results

Frequency of using Homespread Restaurant	Correlation	Sample Size	Significant?*	Strength	Statistical Values t	df	Sig	Outliers Below
With...								0
I am loyal to brands that I like	0.34	1000	Yes	Weak	11.4	998	0.00	0
I have a healthy lifestyle	0.29	1000	Yes	Very weak	9.6	998	0.00	0
I have more ability than most people	0.20	1000	Yes	Very weak	6.4	998	0.00	0
I like to experience things	0.05	1000	No	---	1.6	998	0.10	0
I have an active lifestyle	0.04	1000	No	---	1.3	998	0.20	0
I am an optimistic person	0.04	1000	No	---	1.3	998	0.20	0

Outliers, if any, were removed as per User's request — Outliers:

*Yes=significantly different from zero at 95% level of confidence. Error means no variance or other data error.

Correlation Matrix

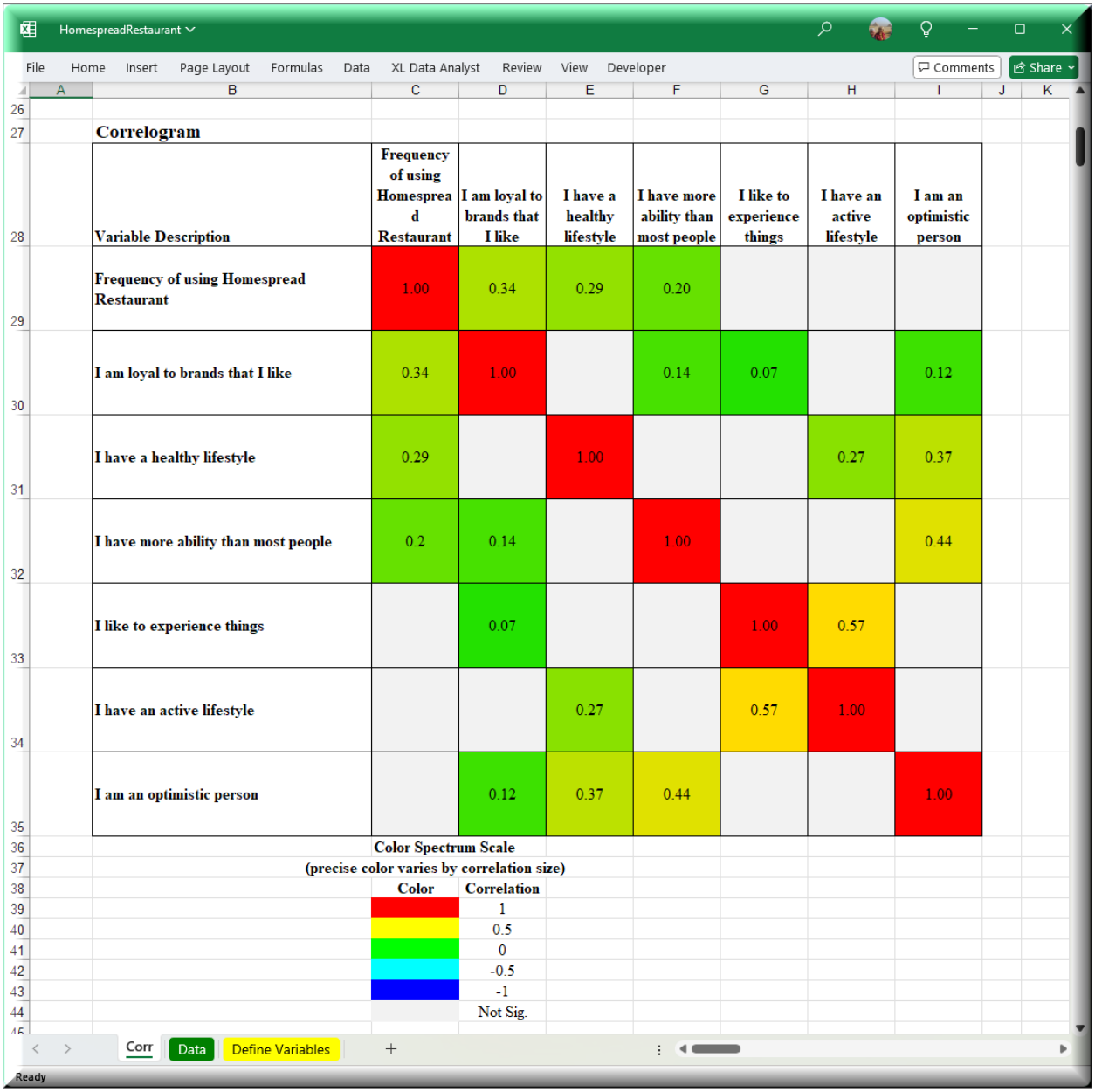
Variable Description	Frequency of using Homespread Restaurant	I am loyal to brands that I like	I have a healthy lifestyle	I have more ability than most people	I like to experience things	I have an active lifestyle	I am an optimistic person
Frequency of using Homespread Restaurant	1.00*	0.34*	0.29*	0.2*	0.05	0.04	0.04
I am loyal to brands that I like		1.00*	0.04	0.14*	0.07*	0.05	0.12*
I have a healthy lifestyle			1.00*	0.01	0.04	0.27*	0.37*
I have more ability than most people				1.00*	0.05	0.02	0.44*
I like to experience things					1.00*	0.57*	0.02
I have an active lifestyle						1.00*	0.02
I am an optimistic person							1.00*

Outliers, if any, were removed as per User's request

*=significantly different from zero at 95% level of confidence. Error means no variance or some other data error.

Correlation Discovery

Discovery – Data Visualization: Data visualization for the correlation matrix is accomplished with a correlogram found immediately beneath the correlation matrix. Only statistically significant correlations are colored with the use of an RGB color spectrum that ranges from red for +1 to blue for -1 with green for a 0 correlation. The precise shade is determined by the correlation's sign and its absolute size.



Correlogram

Variable Description	Frequency of using Homespread Restaurant	I am loyal to brands that I like	I have a healthy lifestyle	I have more ability than most people	I like to experience things	I have an active lifestyle	I am an optimistic person
Frequency of using Homespread Restaurant	1.00	0.34	0.29	0.20			
I am loyal to brands that I like	0.34	1.00		0.14	0.07		0.12
I have a healthy lifestyle	0.29		1.00			0.27	0.37
I have more ability than most people	0.2	0.14		1.00			0.44
I like to experience things		0.07			1.00	0.57	
I have an active lifestyle			0.27		0.57	1.00	
I am an optimistic person		0.12	0.37	0.44			1.00

Color Spectrum Scale
(precise color varies by correlation size)

Color	Correlation
	1
	0.5
	0
	-0.5
	-1
	Not Sig.

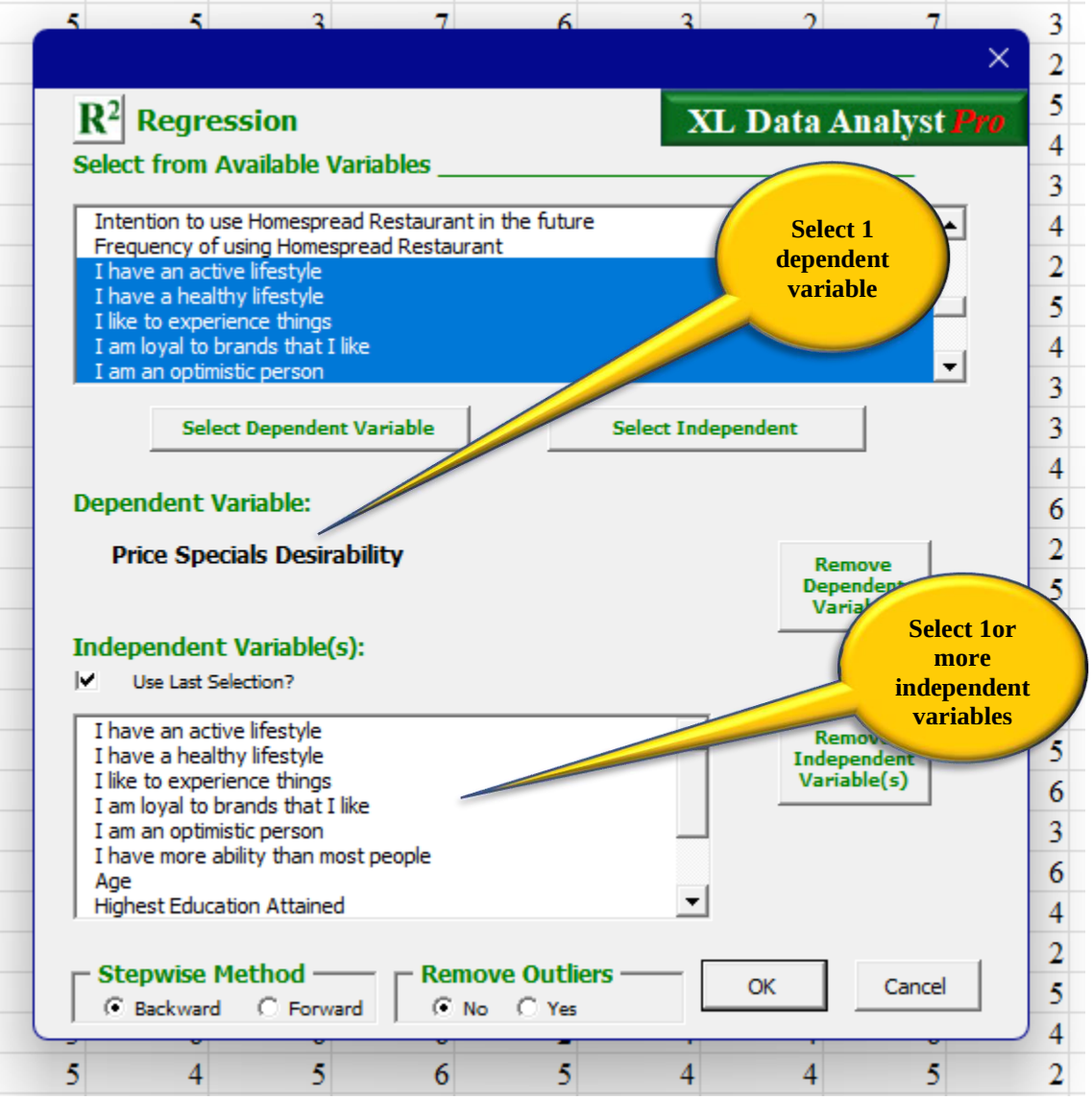
Regression Analysis

Multiple regression analysis involves the use of any number of (mostly) metric independent variables to predict a single dependent metric variable using the following general equation:
$Y = b_1x_1 + b_2x_2 + \ldots b_mx_m$

The analyst selects any number (up to 50, the program limit) of candidate independent variables (x's), and multiple regression procedures identify a reduced final set which are found to be statistically significant (population beta coefficients not equal to zero). The "goodness" of the resulting equation is assessed with Multiple R which is a number ranging from 0 to 1 pertaining to the amount of the variance in the Y variable explained or predicted by the set of significant independent variables using their beta coefficients.

A small number of "dummy" independent variables that are dichotomous categorical (such as males versus female, normally binary coded such as 0,1) can be used; although, diligent interpretation is necessary. *Multiple regression has a great many assumptions, nuances, necessary steps, and options. Users are cautioned to be knowledgeable in these before wholesale reliance on multiple regression findings.*

R^2 **Procedure:** Use the Regression icon to open the Regression variables selection window. With highlighting in the "Select from Available Variables" window, variables are chosen with a button as to "Select Dependent" (single, metric) or "Add Independent" (up to 50, metric and dummy, if desired).



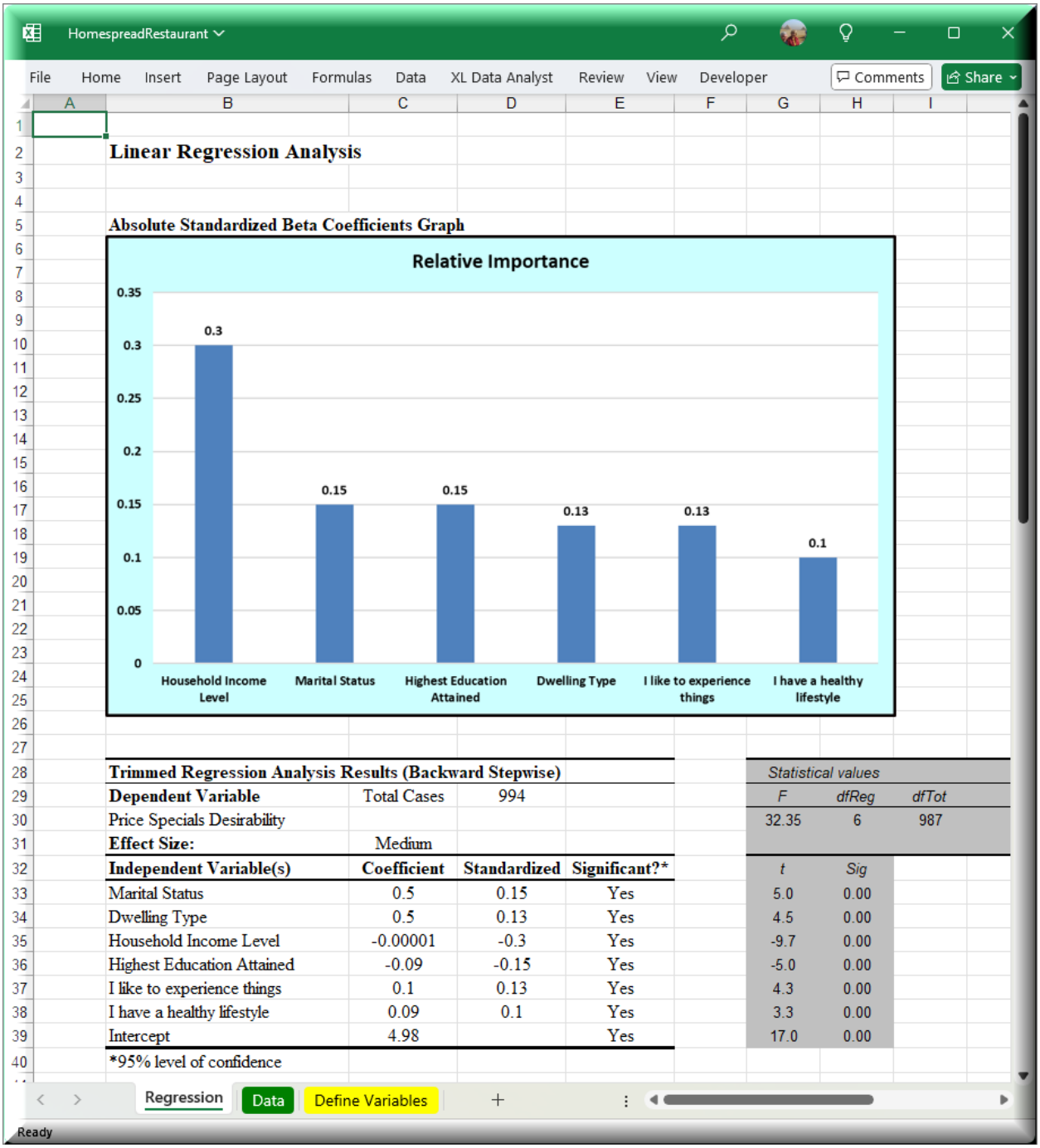
Regression Variables Selection Window

Options: There are three options available on the Regression variables selection window:

- Use Last Selection? Check Box to select the last set of variables analyzed. Default is not checked.
- Remove Outliers? – The user may opt to have any outliers removed from the analysis if found. The default is "No"
- Stepwise Method – The user may select from backward or forward stepwise regression as the method for identifying the final set of all significant independent variables. The default is "Backward."

Discovery – Data Visualization: Data visualization for multiple regression is accomplished in the form of a bar graph presenting the absolute standardized beta coefficients for the statistically significant independent variables determined by the stepwise method chosen by the user. Standardized beta coefficients are often used to judge the relative importance of their respective independent variables in their relationships with the dependent variable.

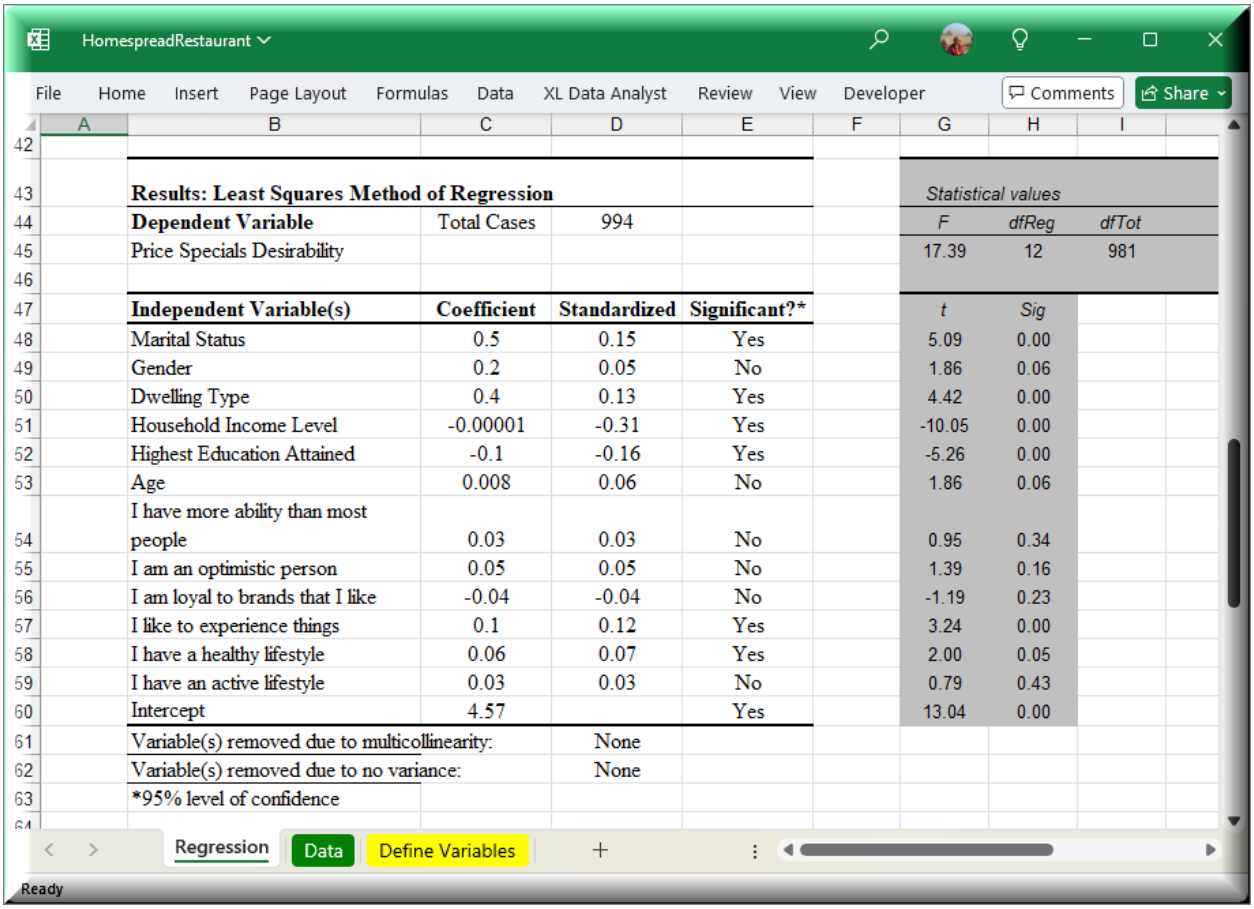
Discovery – Significant Findings: The table immediately beneath the standardized beta coefficients graph is the trimmed result and reports only significant (specified % level of confidence) independent variables based on the chosen stepwise regression procedure. Computed beta coefficients, standardized beta coefficients, and "Significant?*" (all "Yes") are reported. The overall effect size is determined using Cohen's f and reported for the timed, final regression analysis result using the descriptors: "Negligible," "Small," "Medium," or "Large." For advanced users, the accompanying Statistical Values table contains overall model F test results and lists t-test results for each independent variable's beta coefficient.



Regression Data Visualization and Trimmed Results Discovery

Discovery: Initial Findings: The last, or full, multiple regression table contains the results of the entire set of independent variables regressed against the dependent variable. Any independent variable(s) exhibiting no variability or excessive multicollinearity are eliminated during the analysis and identified beneath the table. This table identifies the dependent variable and overall sample size, and it reports the regression analysis with all selected independent variables, their computed beta coefficients as well as standardized betas, and significance finding (Yes or No) for each. For advanced users, the accompanying Statistical Values table contains overall model F test results and lists t-test results for each independent variable's beta coefficient.

If the initial full model F test finds no significant independent variables, analysis stops, and this is indicated. Only the initial findings table can be reported in this case.



42								
43	**Results: Least Squares Method of Regression**					*Statistical values*		
44	**Dependent Variable**	Total Cases	994			*F*	*dfReg*	*dfTot*
45	Price Specials Desirability					17.39	12	981
46								
47	**Independent Variable(s)**	**Coefficient**	**Standardized**	**Significant?***		*t*	*Sig*	
48	Marital Status	0.5	0.15	Yes		5.09	0.00	
49	Gender	0.2	0.05	No		1.86	0.06	
50	Dwelling Type	0.4	0.13	Yes		4.42	0.00	
51	Household Income Level	-0.00001	-0.31	Yes		-10.05	0.00	
52	Highest Education Attained	-0.1	-0.16	Yes		-5.26	0.00	
53	Age	0.008	0.06	No		1.86	0.06	
54	I have more ability than most people	0.03	0.03	No		0.95	0.34	
55	I am an optimistic person	0.05	0.05	No		1.39	0.16	
56	I am loyal to brands that I like	-0.04	-0.04	No		-1.19	0.23	
57	I like to experience things	0.1	0.12	Yes		3.24	0.00	
58	I have a healthy lifestyle	0.06	0.07	Yes		2.00	0.05	
59	I have an active lifestyle	0.03	0.03	No		0.79	0.43	
60	Intercept	4.57		Yes		13.04	0.00	
61	Variable(s) removed due to multicollinearity:		None					
62	Variable(s) removed due to no variance:		None					
63	*95% level of confidence							

Regression Full Results Discovery

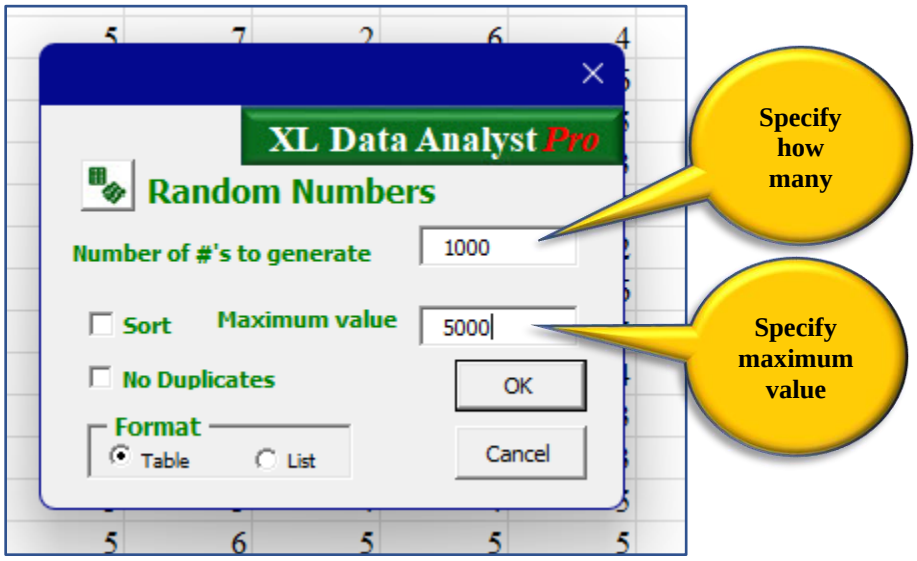
Calculate Features

XL Data Analyst Pro has two calculation options: random numbers and sample size.

Random Numbers

Random numbers are integers that have no systematic relationship with one another. Thus, across a range, such as 1 to 1000, each number is equiprobable, or equally likely to appear. XL Data Analyst Pro will generate a list or a table of random numbers of specified size and within the user's desired range.

Procedure: Use the Random Numbers icon to open the Random #'s window. Using the input window, users may specify a table up to 9,999 random numbers with a maximum value of 999999999.

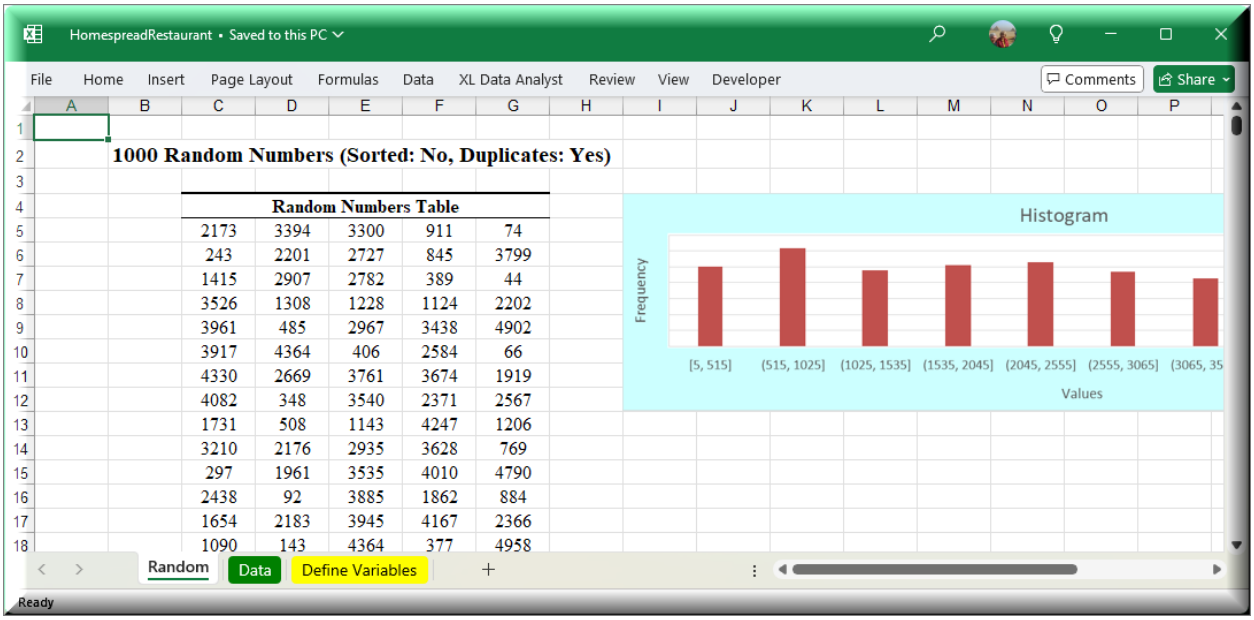


Random Numbers Window

Options: There are three options available with the Random Numbers window.

- Sort Check Box to arrange the random numbers from low to high. Default is not checked.
- No Duplicates Check Box to generate only unique generated random numbers. Default is not checked.
- Format – The user may opt for a table or a list of generated random numbers. The default is "Table."

Discovery: The result is contained on a worksheet named "Random#," and the table or list of random numbers is found on it. The requested number of random numbers is indicated as are the user's specifications with respect to sorting or unique numbers. A data visualization histogram is provided for the user's inspection of the uniformity of the distribution of the random numbers generated.

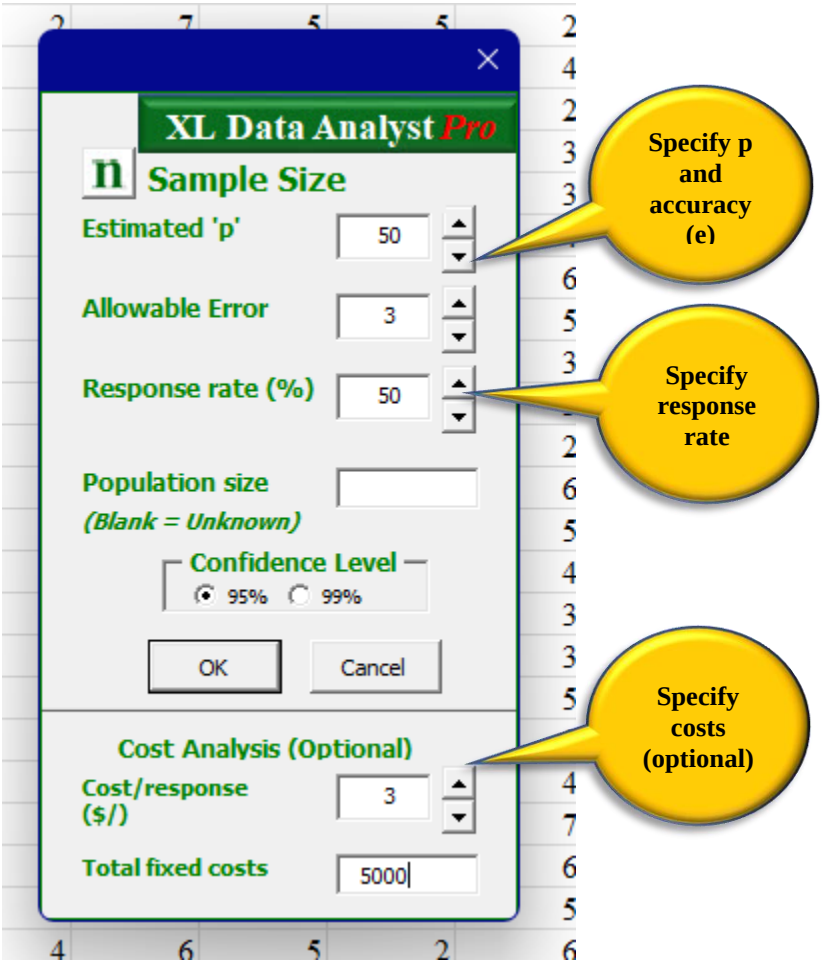


Random Numbers Discovery

Sample Size Calculations

XL Data Analyst Pro will calculate sample size using the standard sample size formula for a percentage at the (e.g.) 95% level of confidence where the user estimates p (percentage), allowable error (e), and population size (optional). The calculation makes the necessary adjustment to take into account an anticipated response rate. In addition, the sample size calculation routine can provide cost analysis information.

Procedure: Use the Sample Size icon to open the Sample Size window. The Sample Size calculation window allows for specification of the "p" value (between 1 and 99) and the allowable error, "e," in whole numbers. The response rate is used to adjust the final sample size. When this Sample Size window opens, the initial values are p = 50%, error = 3%, and response rate = 50%.



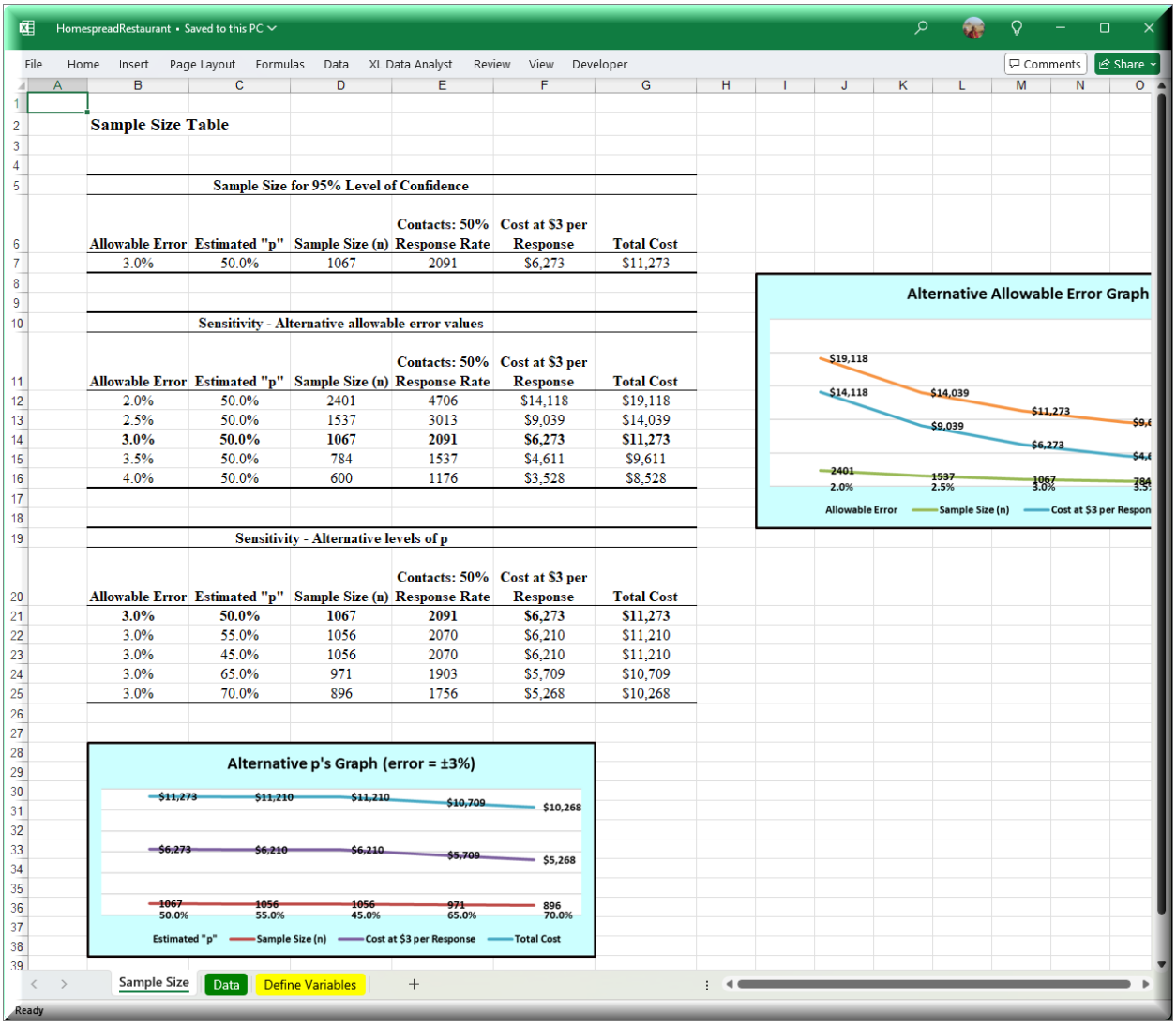
Sample Size Calculation Window

Options: There are four options available with the Sample Size

- Population Size can be entered. The default is blank (unknown).
- Confidence Level – the user may opt for 95% or 99% level of confidence. The default is 95%.
- Cost Analysis data is optional:
  - Cost/response – the user may specify an estimated cost per response. Default is 0.
  - Total fixed costs – the user may specify estimated total fixed costs for the sample plan. Default is 0.

Discovery: The output worksheet is named "Sample Size#," and it consists of three tables. The first table specifies the sample size for the selected level of confidence for the specified p and e values and population size if one is specified. The "Contacts" value is determined by applying the user's estimated response rate. If the user provides cost information, then the table will include a Cost per Response and a Total Cost estimate.

Two other tables contain sensitivity analysis results for: (1) alternative levels of e (e ± .5 and e ± 1.0) at the constant value of p, and (2) alternative levels of p (p ± 5% and p ± 10%) at the constant value of e. These two tables allow the user to consider alternative specifications of p and e without the need for repetitive use of the sample size calculation feature of XL Data Analyst Pro. In these two tables, the row pertaining to the selected "p" and "e" values are in bold for easy identification.



Sample Size Calculation Data Visualization and Discovery

Discovery – Data Visualization: Two graphs are provided with one for the alternative allowable errors table and the other for the alternative levels of p table.

Utilities Features

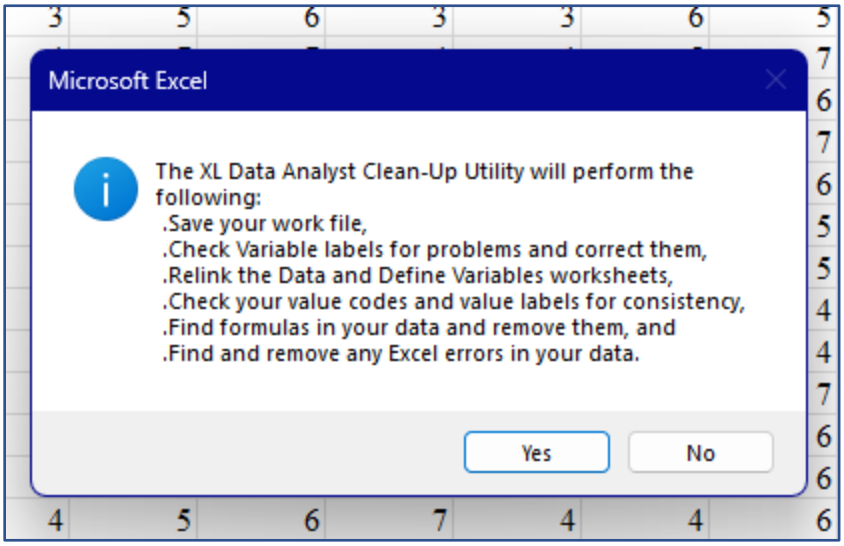
XL Data Analyst Pro has four utilities to assist users in various ways.

Clean-Up

The Clean-Up utility is multipurpose. It is useful to: identify errors in value codes and value labels, find incompatible data (such as formulas), and relink the Data worksheet with the Define Variables worksheet. Whenever users encounter problems with the running of the XL Data Analyst, using Clean-Up may resolve or otherwise identify the sources of these errors.

Users who use their own datasets are strongly advised to run Clean-Up before attempting any analyses as Clean-Up will detect correctable errors that may cause XL Data Analyst Pro to issue warnings or even halt processing.

Procedure: Use the Clean-up icon to initiate Clean-Up.

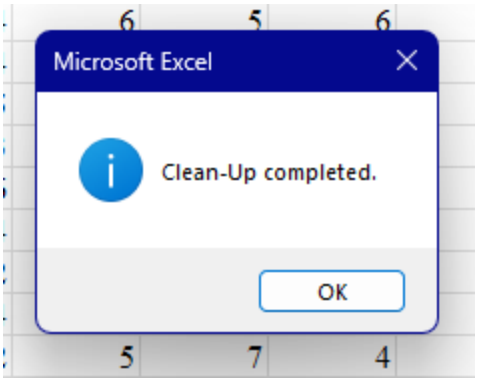


Clean-up Message upon Start-up

As noted in the message box that appears when the Clean-Up utility is activated, Clean-Up performs the following checks and fixes:

- Variable labels – checks for redundant ones and changes the 2nd one to a different label. Eliminates up all punctuation, spaces, and nonalphameric symbols in Variable labels.
- Links the Data sheet to the Define Variables worksheet
- Value Codes and Value Labels – checks that the number of codes is equal to the number of associated labels, and issues warnings if unequal cases are found. The user is prompted to repair the inconsistencies
- Eliminates formulas, if found, in the Data worksheet
- Eliminates errors notations, if found, in the Data worksheet
- Converts any empty dataset cells to blanks (missing data)

Clean-Up takes a few seconds to execute. Upon completion, XL Data Analyst PRO issues a pop-up message that the operation is completed.



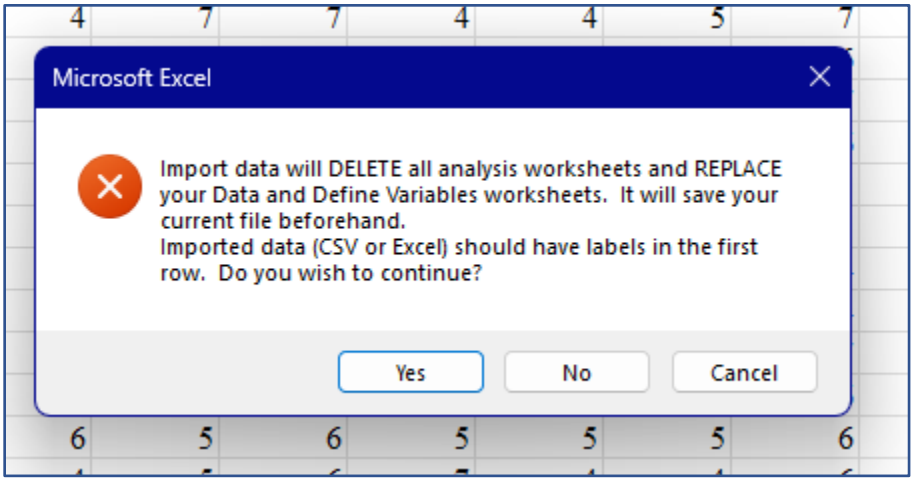
Clean-up Completion Notice

Import Data (Windows only)

The Import Data utility allows users to separately import three types of XL Data Analyst Pro-compatible data files. With comma separated variable (.csv) or Excel data sets users must have a single worksheet (.xls) organized in the rows-by-columns arrangement noted for the Data worksheet, and the first row must have a variable label for each dataset column. With an XLDA (XL Data Analyst) file, the imported file must have "Data" and "Define Variables" worksheets. In all import cases, the imported file dataset completely replaces the XL Data Analyst Pro Data worksheet dataset as well as the "Define Variables" worksheet.

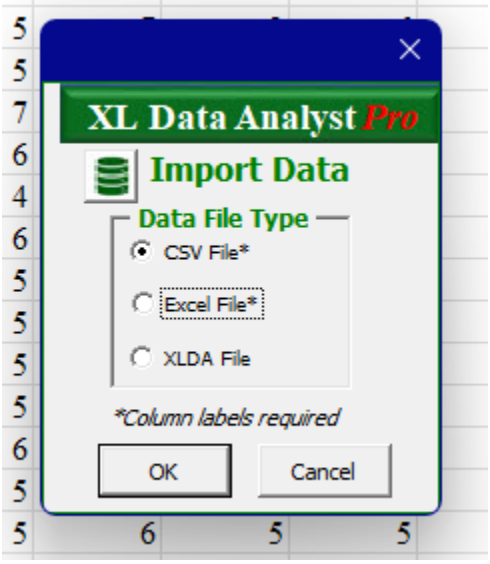
Procedure: Use Data-Import Data icon sequence to open the Import Data window. Users are first reminded that column labels are required for CSV and Excel import files.

A warning is immediately issued to remind the user that all worksheets in the present XL Data Analyst Pro file will be deleted, and the Data and Define Variables worksheets will be modified to accommodate the imported data file.



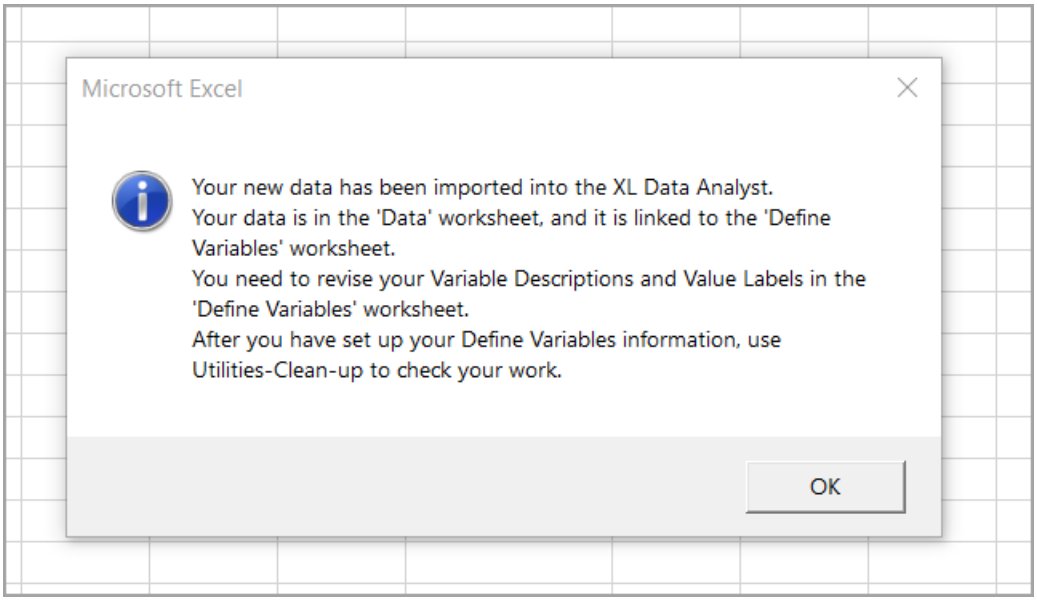
Import Data Warning

The user selects the file type and clicks on OK to initiate the file import process. Instructions are issued for the user to find the file to be imported.



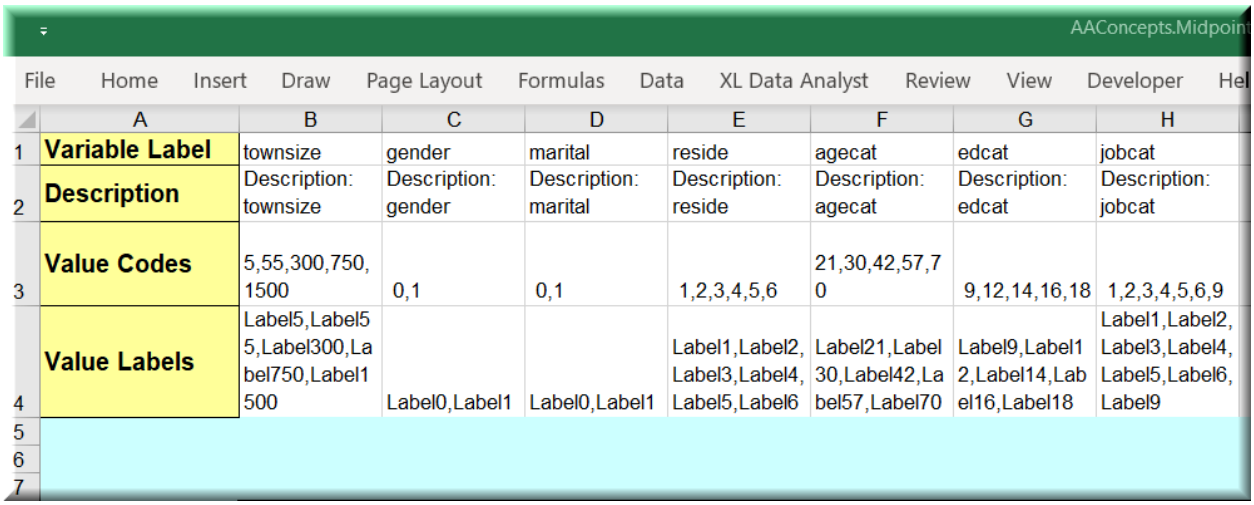
Import Data File Type Selection

At the end of the import, XL Data Analyst Pro issues further instructions.



Import Data Finalization Instructions

In the instance of a CSV or an Excel file import, the Define Variables worksheet is programmed to assist the user in setting up the dataset's value labels. Specifically, the Description for each variable is its Variable name from the Data worksheet, and all eligible unique values are arranged in ascending order, separated by commas, in each variable's Value Codes area. The Variable's Value Labels area is populated with Label#, Label# for each unique value code.



AAConcepts.Midpoint

File Home Insert Draw Page Layout Formulas Data XL Data Analyst Review View Developer Hel

	A	B	C	D	E	F	G	H
1	**Variable Label**	townsize	gender	marital	reside	agecat	edcat	jobcat
2	**Description**	Description: townsize	Description: gender	Description: marital	Description: reside	Description: agecat	Description: edcat	Description: jobcat
3	**Value Codes**	5,55,300,750, 1500	0,1	0,1	1,2,3,4,5,6	21,30,42,57,70	9,12,14,16,18	1,2,3,4,5,6,9
4	**Value Labels**	Label5,Label55,Label300,Label750,Label1500	Label0,Label1	Label0,Label1	Label1,Label2, Label3,Label4, Label5,Label6	Label21,Label30,Label42,Label57,Label70	Label9,Label12,Label14,Label16,Label18	Label1,Label2, Label3,Label4, Label5,Label6, Label9
5								
6								
7								

Define Variables Setup for Import CVS or Excel file

In the case of an XL Data Analyst (XLDA) file import, the Data and Define Variables worksheets are imported wholesale. XLDA import will import either a macro-enabled Excel file (i.e. XLDA Basic file) or an Excel workbook file (see Export Workbook below). If the Data or Define Variables worksheet is not found, a message is issued, and the import function terminates.

Upon completing the import data process and setting up the Define Variables worksheet to pertain to the imported data, users must "Save as…" the file in order to save it under a unique

XL Data Analyst Pro file name. Until the "Save as…" operation takes place, the file will exist under the imported into XL Data Analyst Pro file name.

Here is a summary of what is imported with each file case.

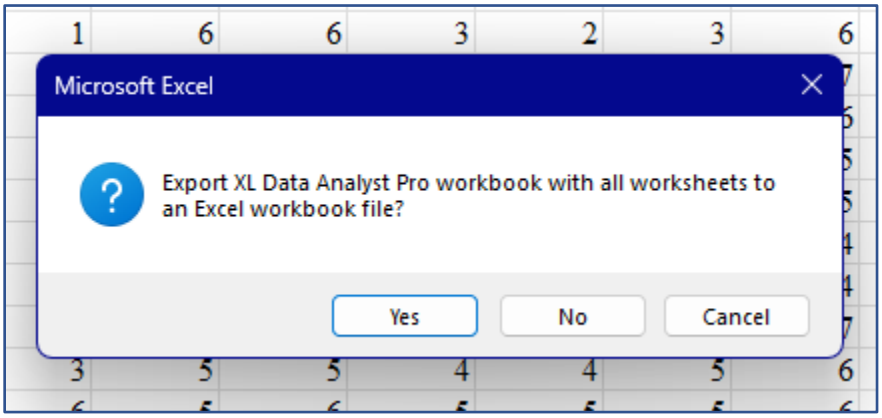
Import Function	Import What	Result
CSV File	Comma separated variable file (single worksheet) with data in columns with variable labels in row 1	Data in Data worksheet, Define Variables populated with Descriptions (column labels), all value codes (1,2,3, etc.) and all value labels (Label1, Label2, Label3, etc.)
Excel File	Excel file (single worksheet) with data in columns with variable labels in row 1	Data in Data worksheet, Define Variables populated with Descriptions (column labels), all value codes (1,2,3, etc.) and all value labels (Label1, Label2, Label3, etc.)
XLDA File	Excel Macro-Enabled workbook file (XLDA file) with Data, Define Variables, and analysis worksheets, if any. Data and Define Variables worksheets are mandatory.	Data and Define Variables worksheets and all other worksheets

Import Data Results

Export Workbook (Windows Only)

The Export Workbook utility allows users to create an Excel workbook file for the current XL Data Analyst Pro file. An XLDA Pro file can be exported to an Excel workbook file format without limitations. The functionality of XL Data Analyst Pro is not included, but users of the .xlxs file(s) can use all worksheets as an Excel workbook file.

Procedure: Use Data-Export Workbook icon sequence. It elicits a dialog box alerting the user that the export function has been activated and asks if the user wishes to proceed. With an affirmative reply, a new Excel (.xlxs, not macro-enabled) workbook is created, and the user is prompted as to where to store the file and what name to assign to it.

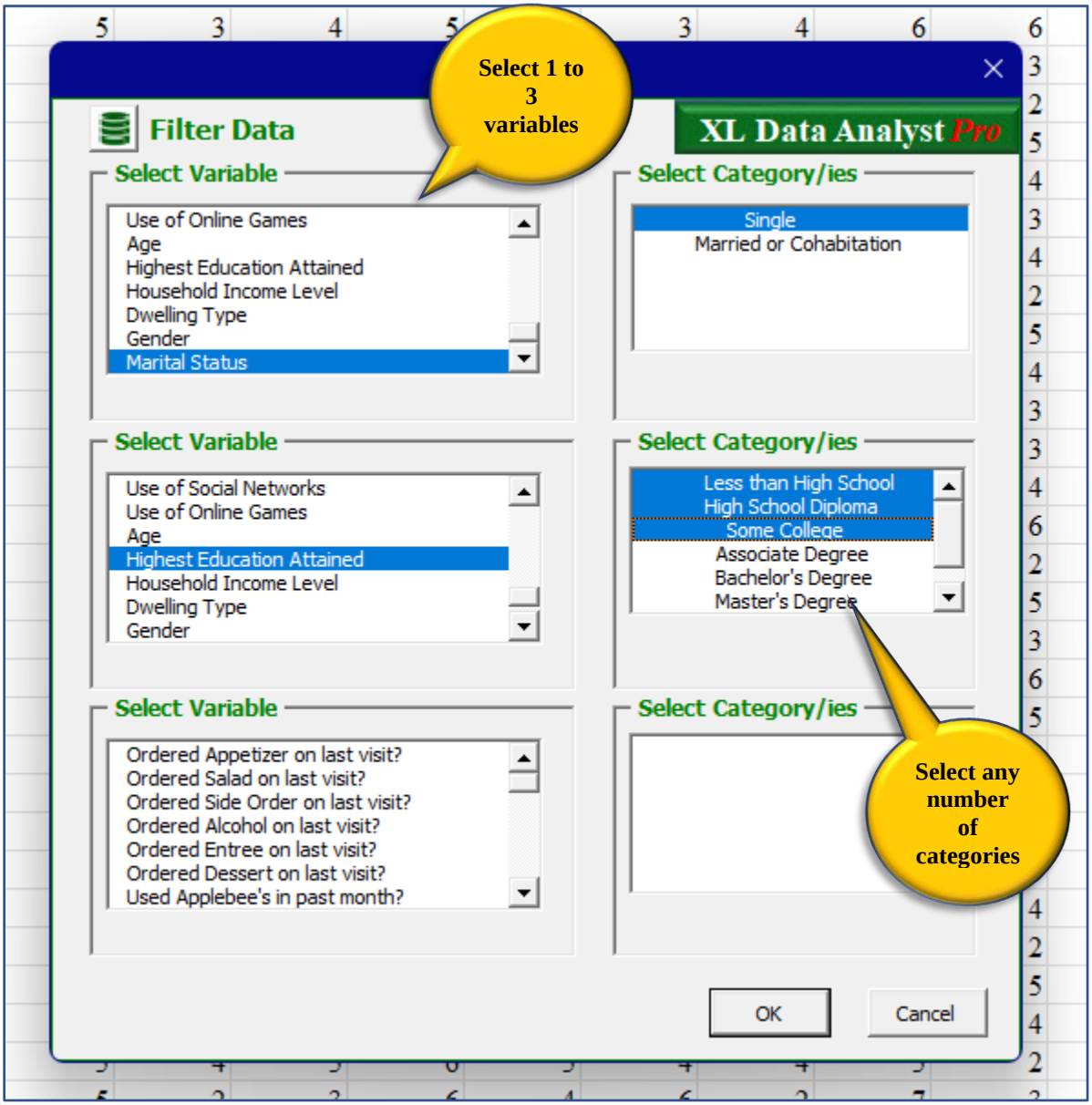


Export Workbook prompt

Filter Data

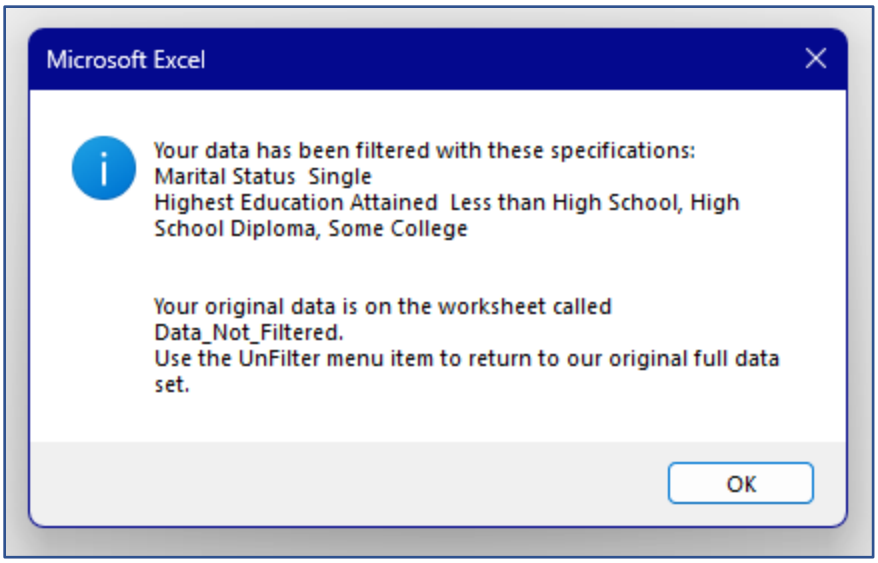
Users who wish to analyze a subset of the full dataset can use the Filter Data utility to identify elements for subsequent analysis. A filtered data set will remain in effect until the user uses the Unfilter Data operation to revert to the original dataset. During analysis of a filtered dataset, XL Data Analyst Pro identifies the filter(s) in place for each analysis.

Procedure: Use the Data-Filter Data icon sequence to open the Filter Data window. Using the cursor to highlight, the user many select up to three different variables for filtering and select any number of categories per variable. All eligible data values for the selected variable appear in the "Select Category/ies" window, not just those that have value labels established.



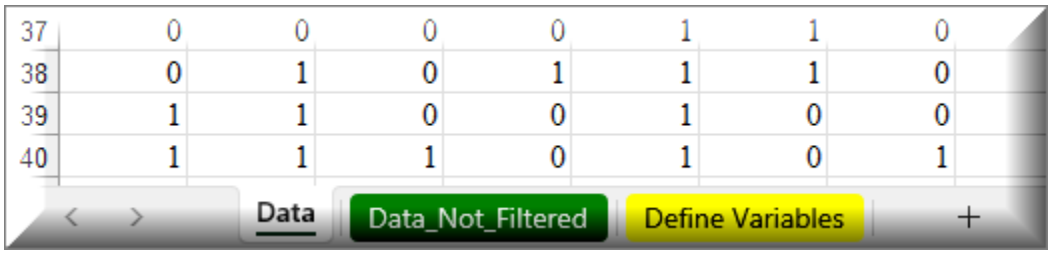
Filter Data Selection Window

After it filters the dataset, XL Data Analyst Pro issues a message as to the filtration system it was directed to implement.

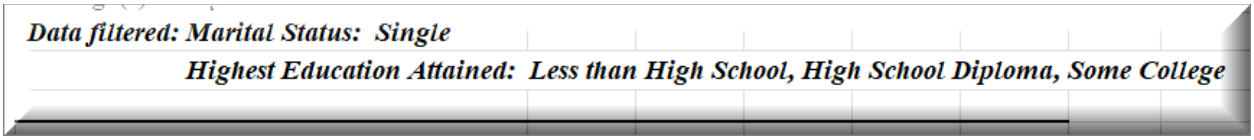


Completed Filter Data Message

With Filter Data, the entire dataset is copied into a new worksheet named "Data_Not_Filtered," and the filtered data is housed in the Data worksheet. Subsequent analyses are performed on the Data dataset, and a note is placed at the top of each analysis specifying the data selection system that is in effect.



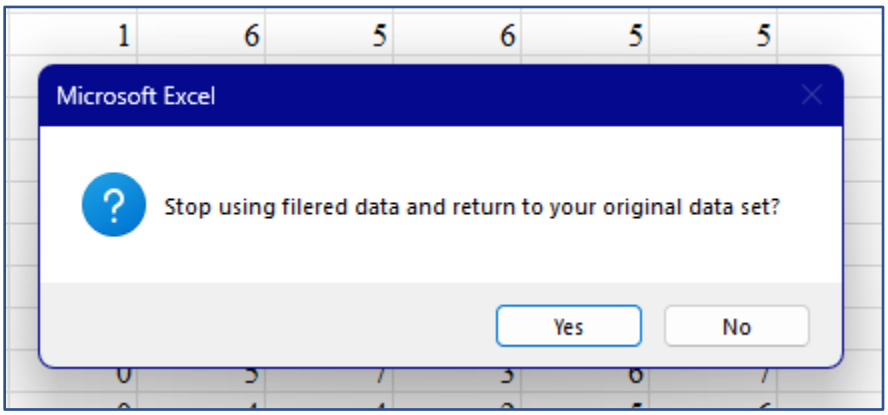
Data and Data-Not_Filtered Worksheets



Filtered Data Notice on Analysis Worksheet

Unfilter Data
The Unfilter Data utility returns the original, full data set.

Procedure: Use Data-Unfilter Data icon to initiate the unfilter data process.



Unfilter Data Message

This operation restores the original dataset to the worksheet named “Data,” and it removes the “Data_Not_Filtered” worksheet. Subsequent XL Data Analyst Pro analyses and operations will take place on the full dataset.

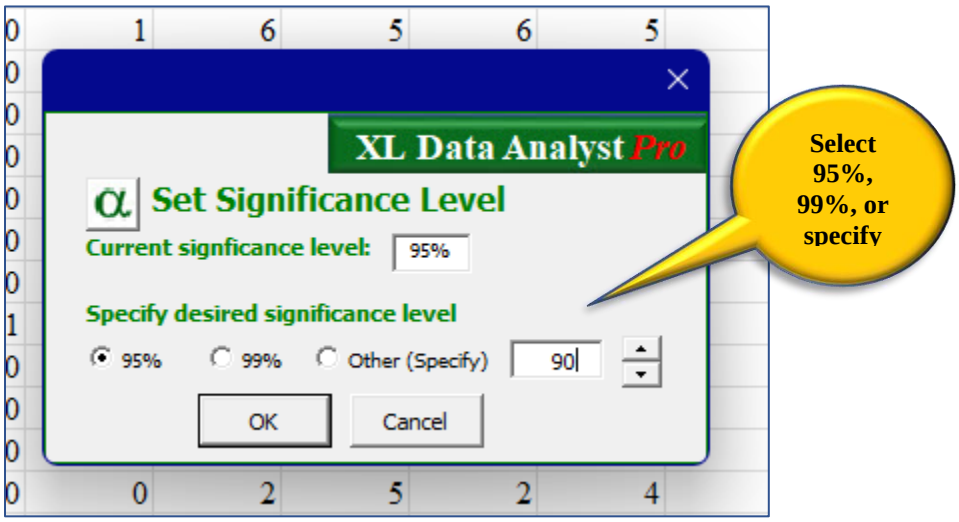
Settings

XL Data Analyst Pro has 5 settings options to allow users to customize its use.

Significance Level

A user may set the level of statistical significance/level of confidence according to his or her risk propensity or for any other reason. Once set, the significance level remains in place until the user resets it with the Significance Level setting or reverts to Default settings. The default is 95% level of confidence.

Procedure: Use the Significance Level icon in the Settings group of the XL Data Analyst Pro ribbon to open the Set Significance Level window. The window identifies the current significance level and allows the user to specify a different one, if desired. The default specification is 95%, and 99% is selectable or the user may opt for "Other (Specify)" and designate a level between 1 and 99 (%). While the "Other (Specify)" option is not selected by default, upon selection, a value of 90% is the initial spinbutton value. For all analyses using significance tests, XL Data Analyst Pro identifies the level of confidence in effect for that analysis.

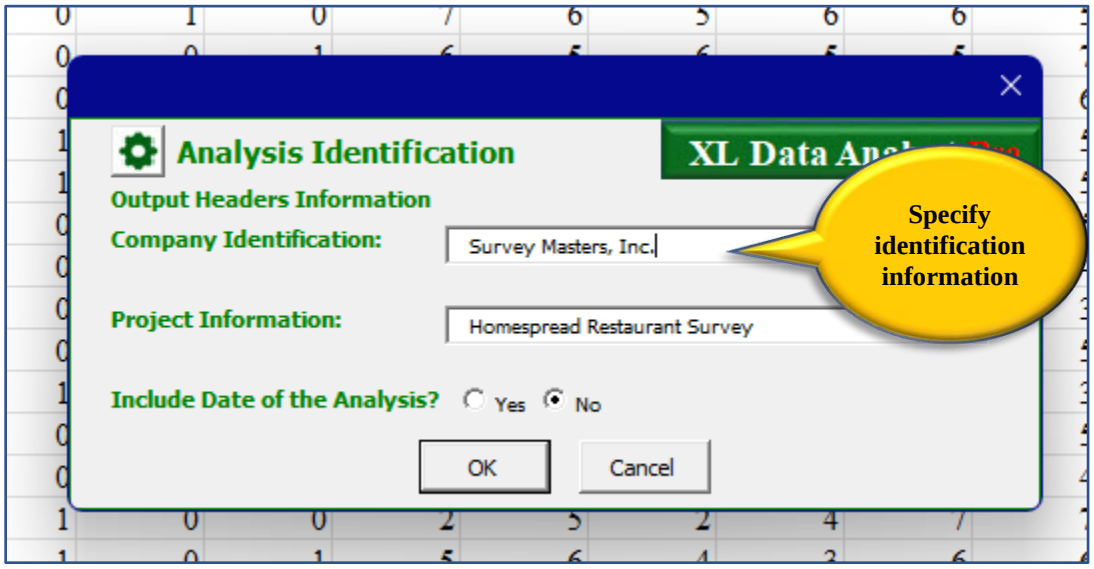


Set Significance Level Window

Identification

There is a setting option for user who wish to specify certain identification information on XL Data Analyst Pro work performed. If specified, this information appears on all subsequent worksheets created by XL Data Analyst Pro. The default is no identification.

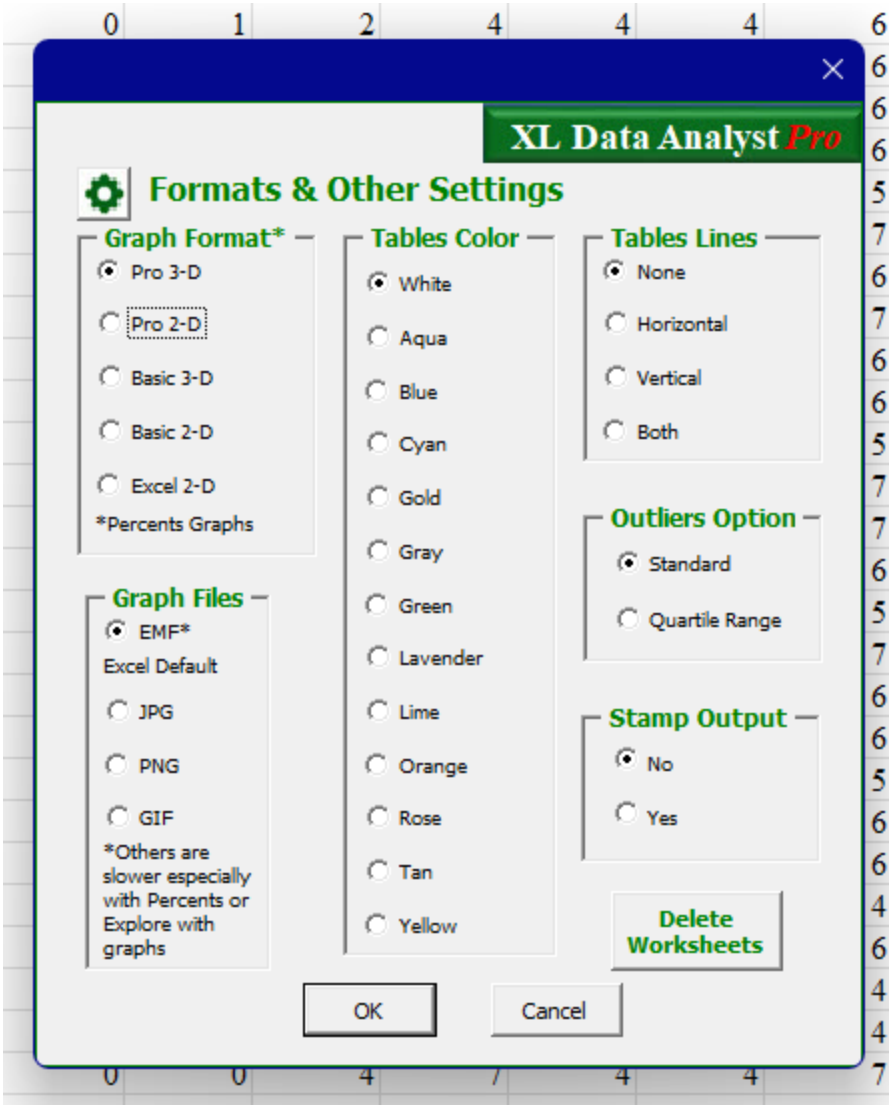
Procedure: Use the Other Settings – Identification icon sequence to open the Analysis Identification Window. Input a Company Identification and/or a Project Information description of up to 50 characters. The labels, "Company Identification" and "Project Information," are arbitrary: the user can place any information he or she desires in these areas. The default is blank, or no identification. If the user desires a date on the output, specify this by clicking on "Yes." The default is no date.



Analysis Identification Window

Format, Etc.
A number of format options are available to XL Data Analyst Pro users: (1) Graph Format for Percents analysis graphs, (2) table lines configuration, (3) color of output tables, (4) XL Data Analysis Pro "Stamp" on or off, (5) method to identify outliers, and (6) delete all analysis worksheets.

Procedure: Use the Other Settings – Formats, Etc. icon sequence to open the Formats & Other Settings window. In the window, use the option buttons to select each format option. Default is the first option of each set.



Formats & Other Settings Window

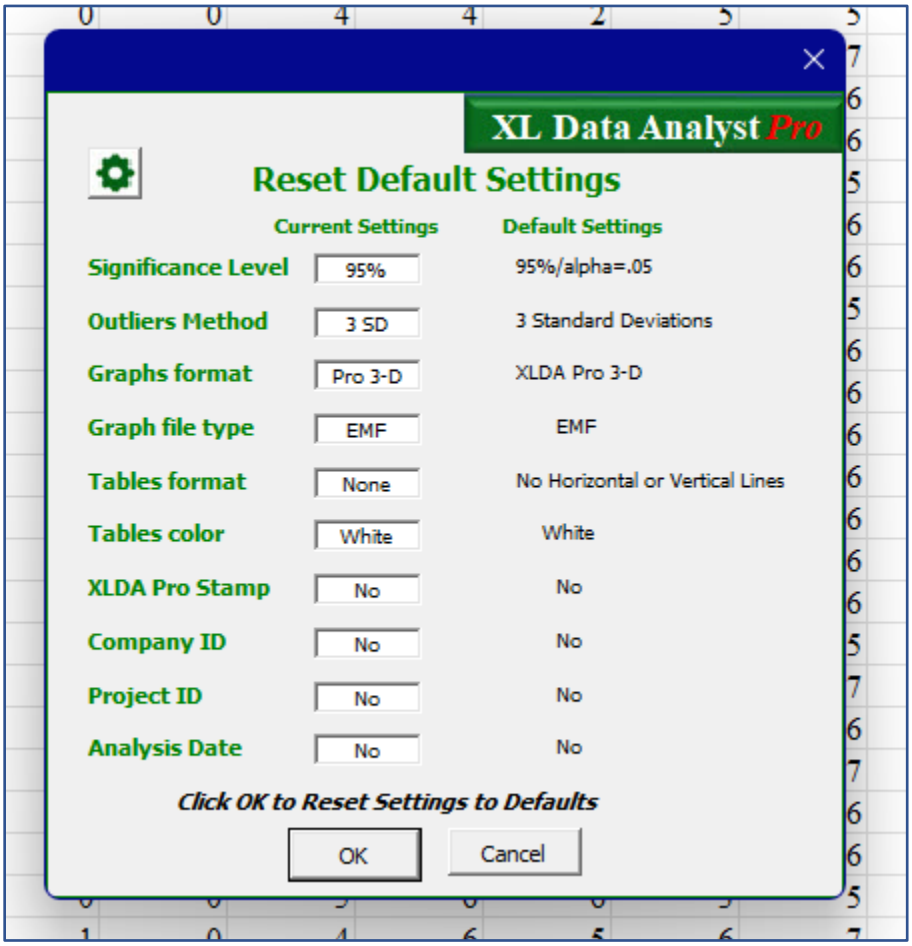
Options: There are five options available in the Formats and Other Settings window.

- Graph Format (for Percents Analysis) – Format options are: Pro 3-D (default), Pro 2-D, Basic 3-D, Basic 2-D, or Excel 2-D
- Tables Color – Color options are: Standard (white, default), Aqua, Blue, Cyan, Gold, Gray, Green, Lavender, Lime, Orange, Rose, Tan, or Yellow
- Tables Lines – Line options are: None (default), Horizontal, Vertical, or Both
- XL Data Analyst Stamp on All Output – Options are No (default) and Yes.
- Outliers Method – Options are Standard Deviation (default) or Quartile Range.

- Delete Worksheets button – deletes all analysis worksheets while retaining Data and Define Variables worksheets and filtered data, if in effect. There is no “undo” for this operation, so users are advised to save-delete-save as to preserve work.

Set Defaults
Users have the ability to restore XL Data Analyst Pro to its ten default settings with a single click.

Procedure: Use the Other Settings-Defaults icon sequence to open the Set Defaults window. The window identifies the Current Settings that are in effect and identifies the Default Settings for XL Data Analyst Pro. A click on the OK button will reset the default settings.



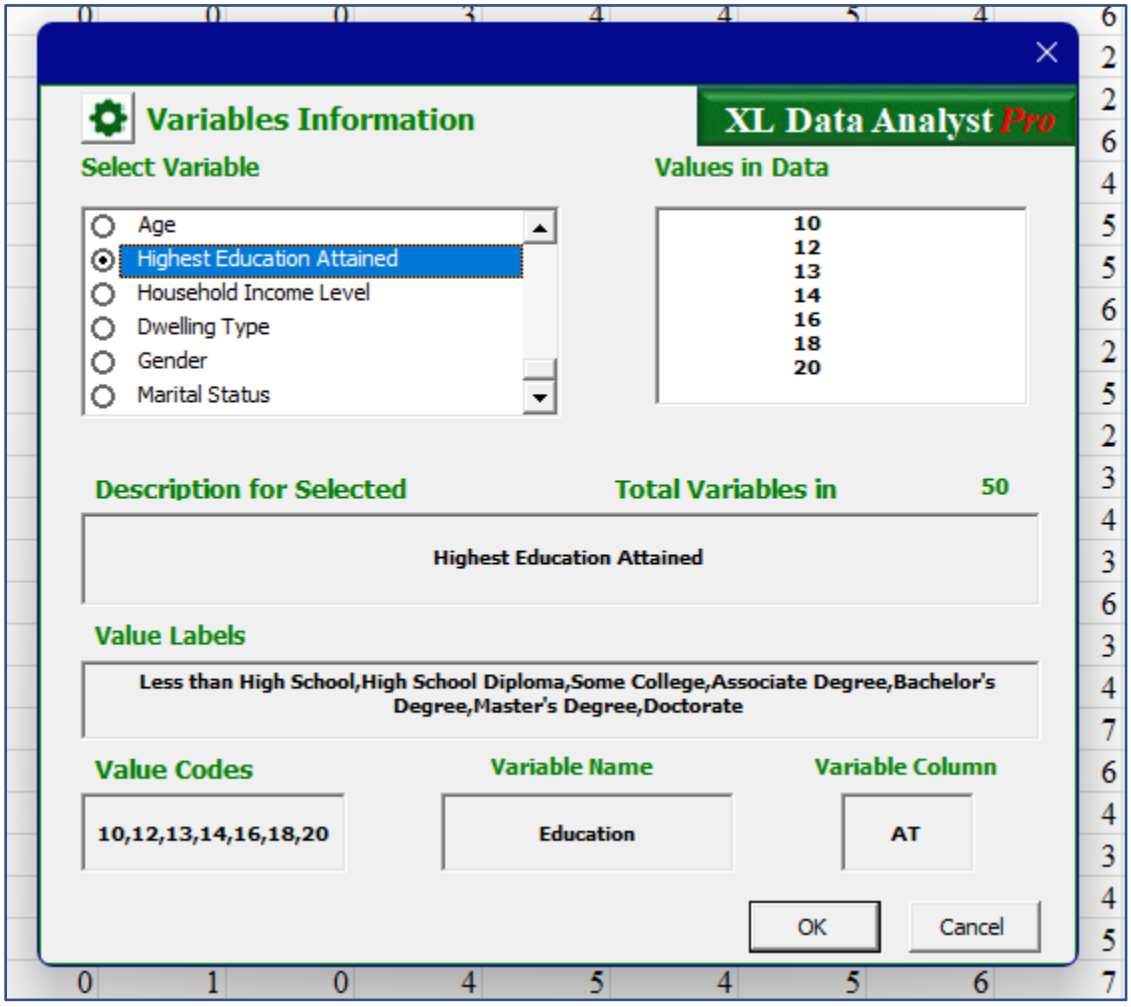
Set Defaults Window

Variables

The Variables option displays the current settings for each variable and allows for reverse coding if appropriate.

Procedure: Use the Other Settings – Variables icons sequence to open the Variables Information window. Select any variable in the "Select Variable" window the following information is displayed:

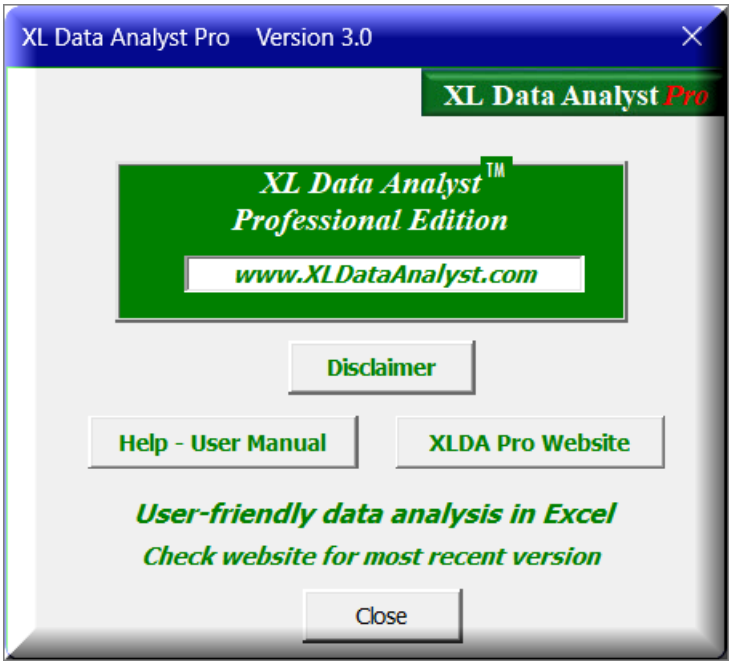
- All unique Values in the Variable's Data
- Full Description for the Selected Variable
- Value Labels (from Define Variables worksheet)
- Value Codes (from Define Variables worksheet)
- Variable Name and Variable column on the Data worksheet



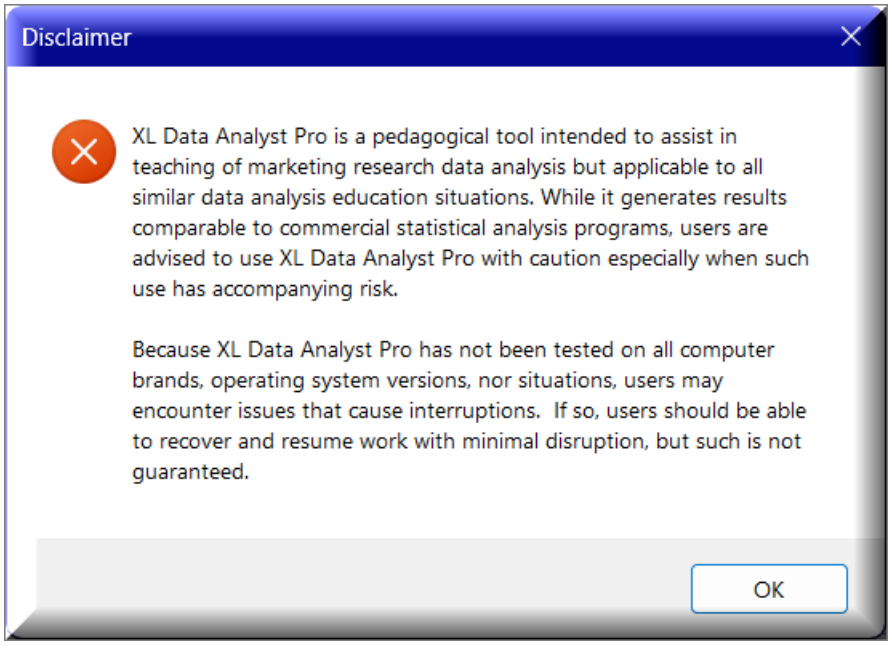
Variables Information Window

About

The XL Data Analyst Pro ribbon About icon opens a window with information about it. Specifically, the About window has links to the XL Data Analyst Pro Manual as well as the XL Data Analyst Pro website where users will find more information and XLDA datasets. In addition, About XL Data Analyst Pro also has a Disclaimer button that issues a message which communicates caution for users of XL Data Analyst Pro.



XL Data Analyst Pro About Window



XL Data Analyst Pro Disclaimer

It is important to remind users that XL Data Analyst Pro had not had extensive testing, and, while its results are comparable to commercial statistical analysis programs, its use apart from a

pedagogical tool is not endorsed. In instances of associated financial, personal, or other risk, users are advised to use commercial statistical analysis programs such as IBM SPSS.

The disclaimer also reminds users that while error catches are programmed into XL Data Analyst Pro, it has not been exhaustively tested, and work may be lost when errors are encountered.

Using the XL Data Analyst with One's Own Dataset

As noted earlier, users can easily apply XL Data Analyst Pro to their own datasets. Recall that there are two worksheets used to set up a dataset in XL Data Analyst Pro. These worksheets are fully established in any XL Data Analyst Pro dataset downloaded from the XL Data Analyst Pro website. Users wishing to apply the XL Data Analyst Pro to their own dataset may do so by correctly substituting their dataset in the Data worksheet and defining their variables, value codes, and value labels in the Define Variables worksheet. When completed, save the dataset under a macro-enabled Excel file name. The XL Data Analyst Pro system will be saved under the new file name.

Setting up the Data Worksheet

Users have two options as to creating a Data worksheet with their datasets: (1) use the Import Data utility, or (2) establish the Data worksheet manually.

1. **Using the Import Data Utility**. This utility will import a comma separated variable (.csv) file or an Excel (.xls) file into the existing Data worksheet of your open XL Data Analyst Pro file. Refer to the description of the "Import Data" utility and notice that users must establish Variable Descriptions, Value Codes, and Value Labels appropriately on the Define Variables worksheet in order to complete the setup of the dataset. Also, until the user saves the imported dataset under a new file name, it exists as the current working file name.

2. **Manual creation of the Data Worksheet.** While this process is more cumbersome, users may opt for manual creation if their own dataset is not amenable to the XL Data Analyst Pro import operation. The steps are as follows:

    a. First, clear the contents of the Data worksheet of any XL Data Analyst Pro file.

    b. Second, if not already present in the raw dataset to be established in XL Data Analyst Pro, create or place Variable labels in the first row of the Data worksheet. The organization of the Data worksheet is in the rows and columns configuration customary to statistical analysis programs. That is, in the case of survey data, the rows correspond to respondents or cases, while the columns pertain to questions on the questionnaire or variables. Each column must have a descriptive label in Row 1.

    Variable Labels in the Data Worksheet. The only requirement for the Data worksheet is to have variable names in Row 1 and numeric or text data in all other rows. Requirements for variable names are as follows.
    - Variable names should be unique, not repeated.
    - Variable names can be any length.
    - Variable names should be letters and/or numbers.
    - Upper- and/or lower-case letters can be used.
    - Variable names should not have spaces in them.
    - The Labels function of Excel does <u>not</u> need to be in effect.

    a. Input Data into the Data Worksheet. Since Row #1 contains the variable names, the dataset raw data necessarily should reside in Row #2 to the last row that constitutes

the dataset. Users whose datasets are in spreadsheet organization may accomplish this with a copy and paste operation.

Once the Data worksheet has been established with the user's own dataset, it is prudent to save the work as an Excel file in the user's dataset name.

Defining Variables in the Define Variables Worksheet

The Define Variables worksheet contains Variable Labels, Descriptions, Value Codes, and Value Labels.

If the user has used the Import Data utility, the prior XL Data Analyst Pro Define Variables contents will be replaced as noted in the "Import Data" utility description. The new Variable Labels will be on the Define Variables worksheet in the Descriptions cells, and the Data and Define Variables worksheets will be linked properly. If this is the case, skip to "Variable Description" below.

Users who use the manual creation option must clear the contents of the XL Data Analyst Pro file being modified for their own datasets. To clear the contents block B1:B4-to-the-end of the Define Variables current dataset variables definitions and use "Clear Contents" or Delete.

Under each variable name in the Define variables, enter in the Variable Description, Value Codes, and Value Labels according to the following instructions.

Variable Descriptions. In Row 2, type or paste in long descriptions of the variables. These descriptions will appear in XL Data Analyst Pro tables. There is no limit to the length and any letter, number, punctuation mark or symbol is acceptable. Complete variable descriptions will show completely in the Define Variables worksheet as cells have been formatted with Format Cells – Alignment – Wrap Text to make the complete descriptions appear.

Value Codes. Where the variables are coded with numbers that pertain to groups or categories (e.g. 1=male, 2=female; 1=married, 2=single, 3=single survivor, etc.) enter in the code numbers, each separated by a comma. Note: It is vital that users **do not use spaces before or after the commas in the value codes** as the XL Data Analyst Pro takes these into account. For example, if the user specifies 1, 2, then the XL Data Analyst Pro will interpret the codes as "1" and " 2" meaning that a "2" in the Data worksheet for that associated variable will be treated as different from a " 2"code. Also, Clean-Up will not detect spaces in the value codes.

Value Labels. In the cells directly under the cells where value codes have been entered, enter the corresponding value labels, each separated by a comma. Again, it is vital that users **do not use spaces before/after commas in the value labels** as the XL Data Analyst Pro takes these into account. Value labels will appear in XL Data Analyst Pro tables in place of the value codes.

When setting up Value Codes and Value Labels, it is important to remember that a comma is used to denote the separation of one code or label from another, and the XL Data Analyst expects the listing order of the labels to be parallel that of the codes. Thus, codes "1,2,3" and labels "red,blue,green" are taken as 1 for red, 2 for blue, and 3 for green. Note that spaces are taken literally. In other words, "1, 2, 3" indicates data codes of "1", " 2", and " 3", and "red, blue, green" indicates labels of 'red", " blue", and " green". It is advisable to use no spaces before or after data codes on the Define Variables worksheet, while spaces in data labels are discretionary. In the case of a metric variable, it is acceptable to not have any value codes or value labels.

Where variables are metric, that is, natural numbers such as age, number of times, dollar spent, etc., value codes and value labels are not entered. Simply leave the Value Codes and Value Labels cells for that variable blank.

Linking the Data and Define Variables Worksheets

To link the Data worksheet to the Define Variables worksheet, use Utilities-Clean Up. Alternatively, select all labels in Row 1 on the Data worksheet and Paste-Special, Paste Link to cell B1 on the Define Variables sheet.

Using the Clean-Up utility to check the integrity of the Define Variables worksheet with respect to matching up with the data in the Data worksheet as it will note any discrepancy in the number of value codes vis-à-vis value labels for any variable.

Also, the user is reminded that the new file must be saved with a "Save as…" Excel operation in order to be established as a separate XL Data Analyst Pro macro-enabled (.xlsm) file with a unique file name.

Troubleshooting

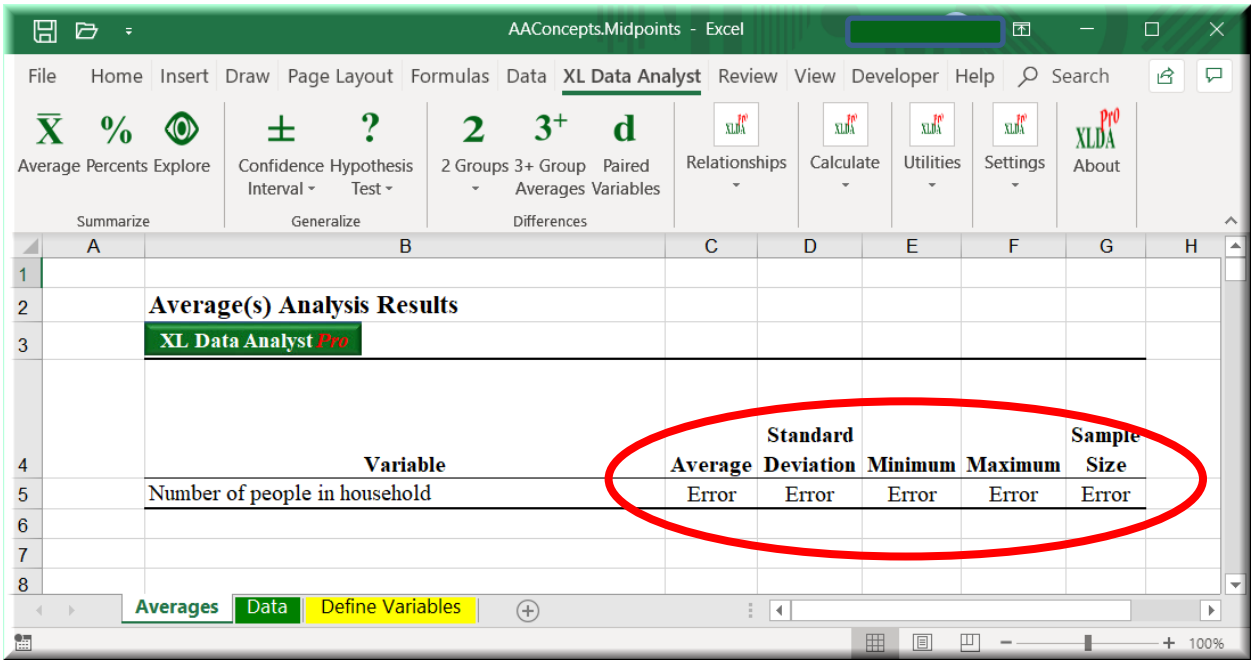
Note on Testing of XL Data Analyst Pro

As noted, XL Data Analyst Pro has had extensive testing in a Windows environment. However, this testing has been exclusively on Dell pc's or laptops, so problems with other computer brands are unknown. As for Mac testing, it has been much less extensive and limited to MacBook Pro models. Because XL Data Analyst Pro was only recently converted to run on Macs, Mac users may encounter problems that are not addressed here.

Warnings built into XL Data Analyst Pro

Excel will issue errors if it is required to perform operations such as division by zero. XL Data Analyst Pro has been programmed to inspect data involved with its analyses and to warn users with a pop-up message. Typically, processing continues; however, some errors result in termination of the analysis underway.

With errors that do not terminate processing, the XL Data Analyst Pro will indicate "Error" or "N.A." on its output tables so users may determine the offending variable(s). See the following example.

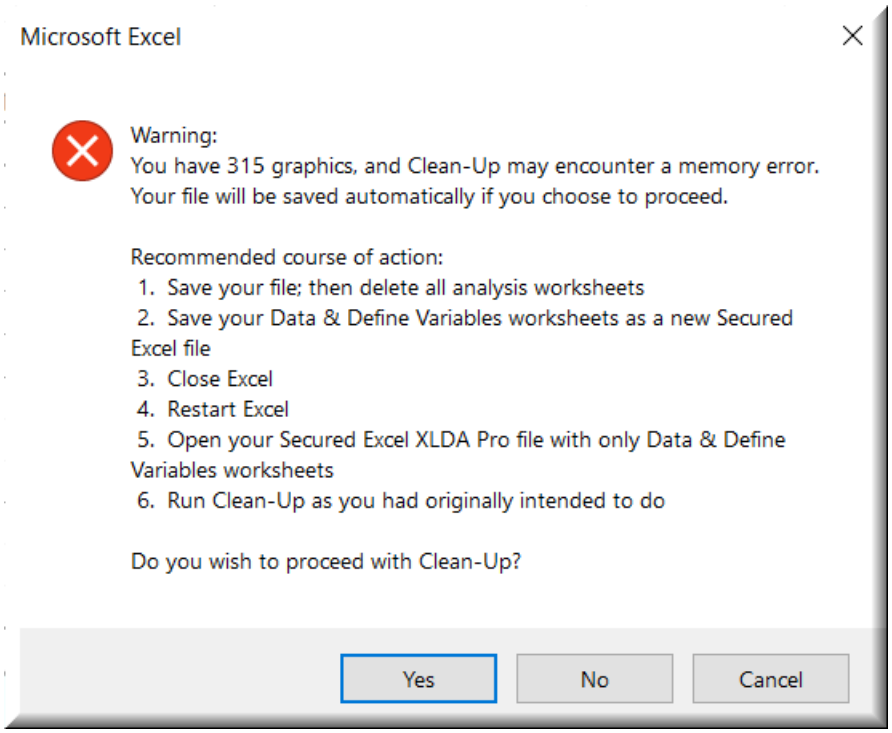


XLDA Data Error in Discovery Table

With respect to files downloaded from the XL Data Analyst Pro website, these files should not have errors. It is possible for users to encounter errors if they inadvertently change the Data worksheet or the Define Variables worksheet. If an error that causes a Visual Basic error message (see below) occurs, users are advised to re-download the file.

Out of Memory Possible Warnings

Because XL Data Analyst Pro creates high resolution graphs, depending on the user's pc memory size, heavy graphing sessions may encounter out-of-memory errors. Both Clean-Up and Import Data issue warnings if more than 100 graphics are in the current workbook. Additionally, Explore with graphs generates 3 graphs per variable analyzed, and low memory errors may occur during this analysis. In these three cases, the XL Data Analyst Pro saves the current file and recommends the following steps for resolution.

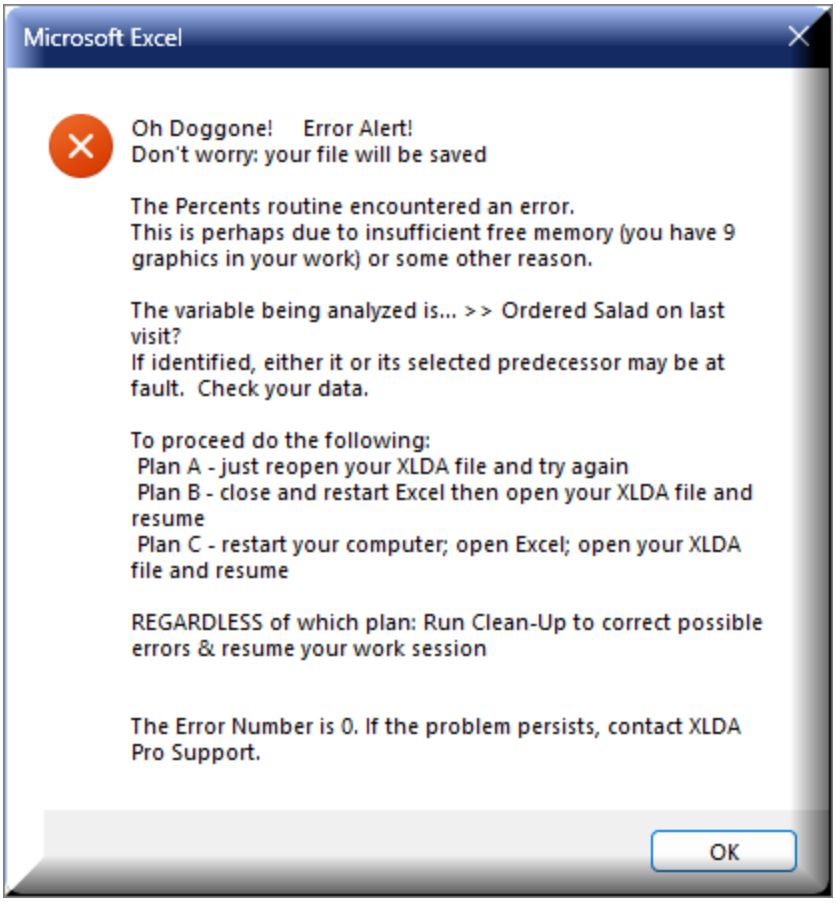


Out-of-Memory Possible Warning Message

XL Data Analyst Pro Fatal Error Catch

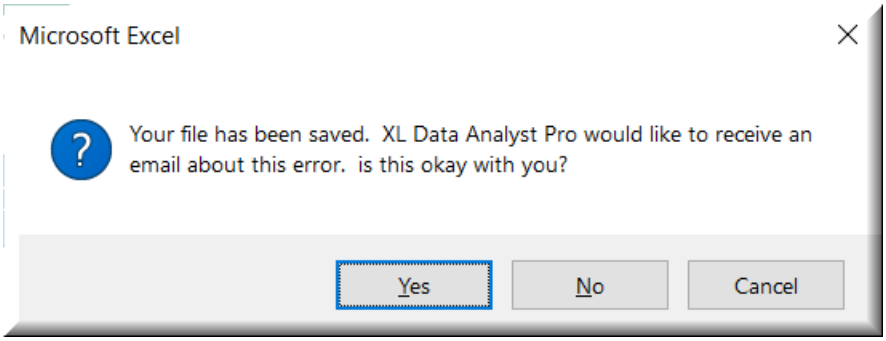
Within each module and function of XL Data Analyst Pro is built an error catching mechanism that avoids crash-and-exit when a major error occurs. Normally, this is a case of too many graphs which take up computer memory. If this or any other error disrupts processing, XL Data Analyst Pro will issue an error alert such as follows.

This error catch mechanism identifies the module being processed, if possible, and the variable under analysis if one is the case. The error catch describes steps to recovery, including an automatic save of the current work file. Error catch saves the current work file and closes Excel. Note that recovering one's work may require a computer restart (See Plans A, B, C in the message).



XL Data Analyst Pro Major Error Message

As part of its quality control program, XL Data Analyst Pro wishes to receive an email record of such errors encountered by users. Consequently, after the error catch mechanism issues its message, the user will be prompted about this email message. If the user agrees, the email message will be sent via Outlook if it is installed on the user's computer. This email message is automatic, and the user is not required to provide any description or text.



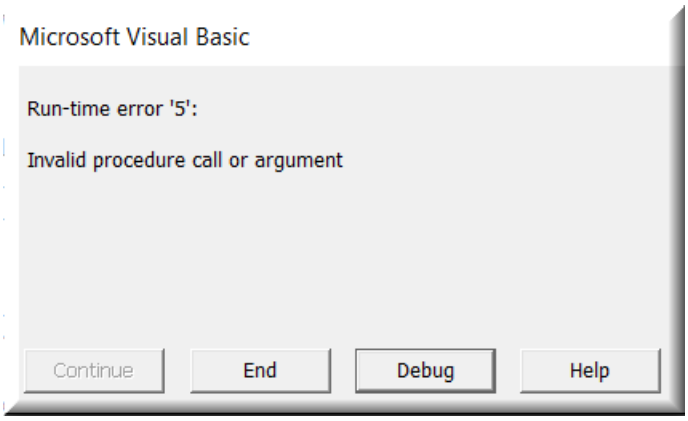
XL Data Analyst Pro Email Notice

Visual Basic Errors

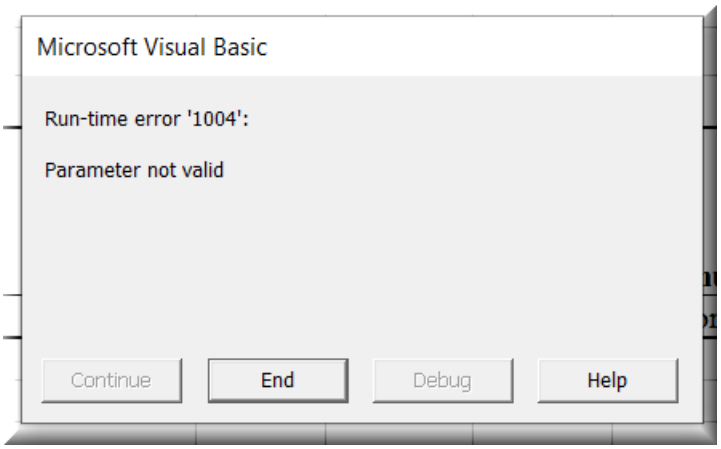
While exceptionally rare, users who apply XL Data Analyst Pro to their own data or who intentionally change the downloaded original XL Data Analyst Pro files may encounter Visual Basic errors. Specifically, errors that are internal to the XL Data Analyst Pro (i.e. Visual Basic errors) will be encountered if a user does not adhere to the requirements for Excel and/or the XL

Data Analyst Pro. Note: While XL Data Analyst Pro has been extensively tested with Windows, considerably less rigorous testing has been done with Macs, so VB errors may occur.

Below are Visual Basic error messages, which may appear differently on Macs, and likely causes/solutions. Generally, these errors are intercepted and the Fatal Error Catch message above is issued instead of the Visual Basic error message.



Visual Basic Error Message



Visual Basic Error Message

Possible Causes: On the Define Variables worksheet, the number of value labels is different from its corresponding number of value codes for one or more variables. Or the variable labels on the Data worksheet are not properly linked to the Define Variables. Alternatively, there may be a data error causing division by zero, Excel errors, or some other computation error.

Solution: Run Clean-up to check that the number of value codes is equal to the number of value labels for each variable. Clean-Up will relink the Data and Define Variables worksheets. Clean-up will also correct data errors. If the error persists, inspect all of the data for the offending variable to ensure it is numeric or otherwise not consistent with Excel requirements.

This manual was prepared by

Alvin C. Burns

Head & Professor Emeritus/Retired
Department of Marketing
E.J. Ourso College of Business
Louisiana State University
Baton Rouge, Louisiana

Correspondence:
alburns@lsu.edu or alcburns@gmail.com