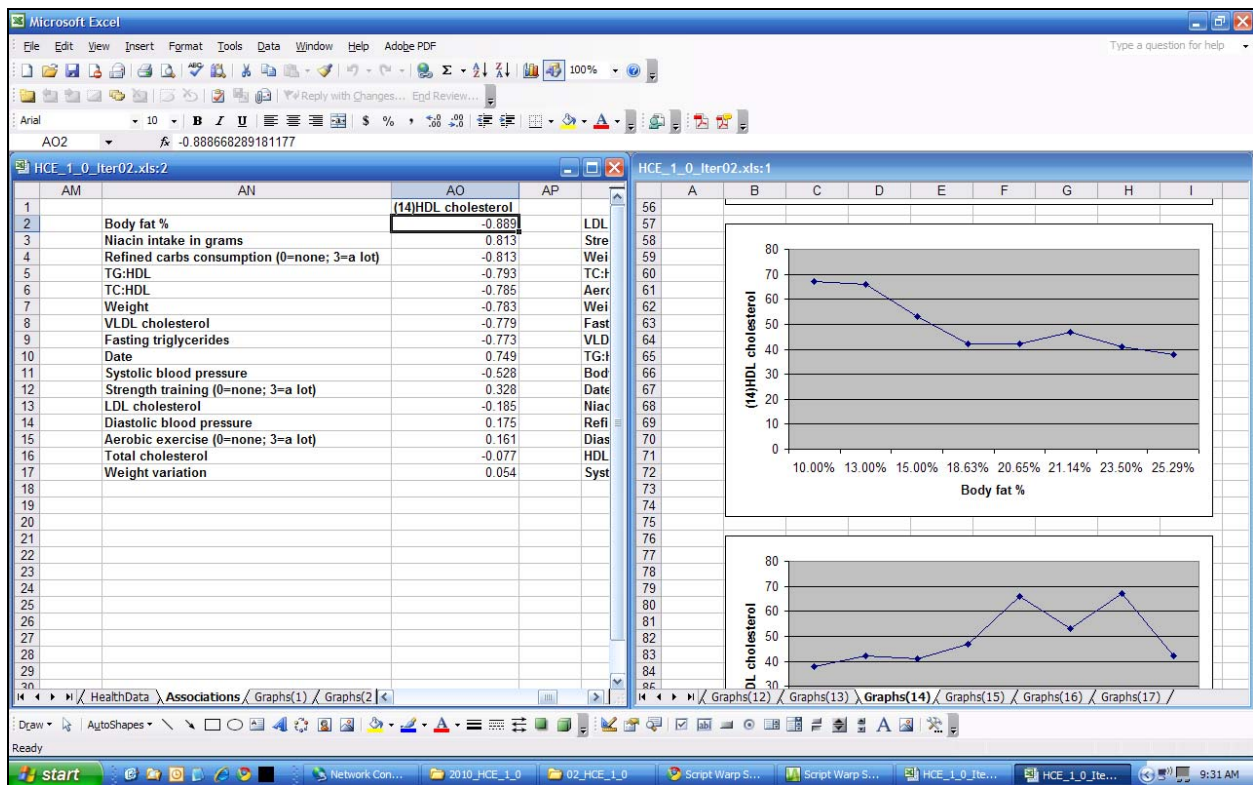# HealthCorrelator for Excel 1.0

User Manual

Ned Kock

HealthCorrelator© for Excel 1.0 User Manual

Ned Kock

ScriptWarp Systems ™
Laredo, Texas
USA

HealthCorrelator for Excel 1.0 User Manual, November 2010, Copyright © by Ned Kock

Software use agreement

The use of the software that is the subject of this manual (Sofware) requires a valid license. Individual and organizational licenses may be purchased from ScriptWarp Systems, or any authorized ScriptWarp Systems reseller.

The Software is provided "as is", and without any warranty of any kind. Free trial versions of the Software are made available by ScriptWarp Systems with the goal of allowing users to assess, for a limited time or number of use sessions, the usefulness of the Software for their data modeling and analysis purposes. Users are strongly advised to take advantage of those free trial versions, and ensure that the Software meets their needs before purchasing a license.

It is strongly recommended that the Software be used in conjunction with advice from a licensed health care professional. The Software is designed to be used as a self-help tool, based exclusively on the instructions in this document. The Software's developer may provide suggestions and tips through various sources (e.g., blog, web-based video clips). However the developer does not guarantee help, support, or maintenance of the Software after it is purchased.

Statistical analysis software systems are inherently complex, sometimes yielding results that are biased and disconnected with the reality of the phenomena being modeled. Users are strongly cautioned against accepting the results provided by the Software without double-checking those results against: past empirical results obtained by other means and/or with other software, applicable theoretical models, and practical commonsense assumptions.

Under no circumstances is ScriptWarp Systems to be held liable for any damages caused by the use of the Software. ScriptWarp Systems does not guarantee in any way that the Software will meet the needs of its users.

For more information:

ScriptWarp Systems
P.O. Box 452428
Laredo, Texas, 78045
USA

www.scriptwarp.com

Table of contents

1. What is HCE?

HealthCorrelator for Excel (HCE) is a basic health data analysis application that runs on Microsoft Excel. It generates: (a) linear regression coefficients of association among health variables; (b) lists ordered by those coefficients of association; and (c) graphs plotting linear or nonlinear relationships among health variables.

No two individuals respond in exactly the same way to health interventions, whether those interventions are based on diet changes, lifestyle changes, drugs, or other factors. HCE can be used to identify health-related cause-and-effect relationships for a patient, and the nature of those relationships, thus allowing doctors and patients to develop tailored health interventions.
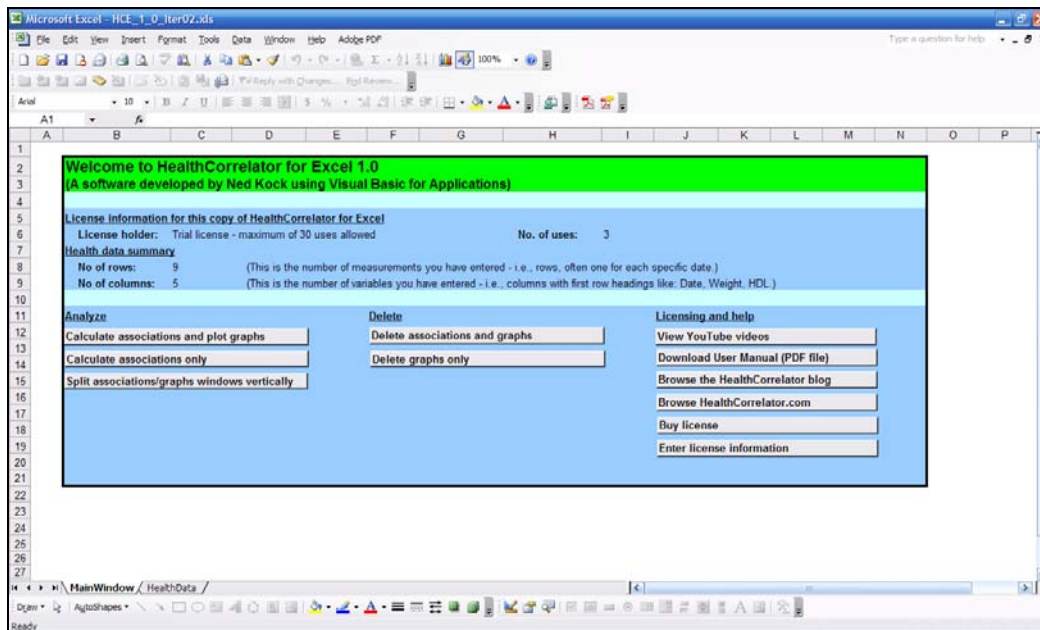
HCE can also be used by athletes and trainers to identify the strength and nature of relationships between performance measures and various causative factors, particularly dietary and training-related factors. As with health-related improvement, HCE can be effectively used for athletic performance improvement, by providing information to athletes and trainers for tailored interventions.

2. Software installation and uninstallation

HCE is made available as an Excel file with embedded "macro" code. It is designed to run on any computer with a Windows operating system, as long as Excel 2003 or a newer version of this spreadsheet software is installed.

To install HCE, simply download it to the desktop or to a folder, and click on the file. It will be automatically opened with Excel, as long as Excel is also installed on the computer. **Macros must be enabled in Excel for the software to run.** The embedded code will install HCE and show you the main window of the software with several options (see Figure 2.1). To uninstall HCE, simply delete all copies of the software that were either downloaded to your computer or created by you through copy-and-paste.

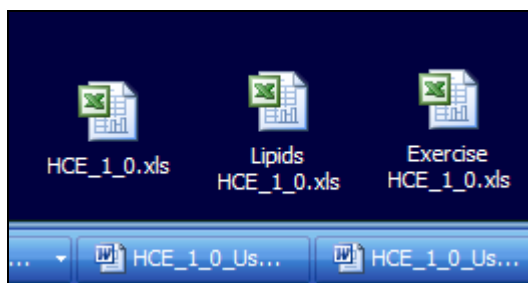**Figure 2.1. First window shown after installation**

At first install, the trial license will be made available for 30 "uses"; that is, 30 use sessions. During or after this period a user can buy a license. After buying a license, the user will receive a license key. After clicking on the button to enter license information, the user will be asked for the license key and additional information.

3. Making copies of HCE to analyze multiple health datasets

The software comprises several sheets, which are part of one main workbook. The various sheets in each workbook refer to analyzes of a single health dataset. Given this, **it is strongly recommended that users create separate copies of HCE, each with the main health dataset that they want to analyze.** The original copy of HCE, which comes with a sample dataset, should be kept intact as a basis for creating other copies.

On Figure 3.1, the original copy of HCE is shown on the left, with two copies to the right created for the analysis of lipid panel and exercise data. These are saved on a computer's desktop folder. Users are welcome to rename the copies as they see fit, but it is recommended that they keep the "HCE_1_0" part so that they know that each copy refers to a particular version of HCE (in this case version 1.0).

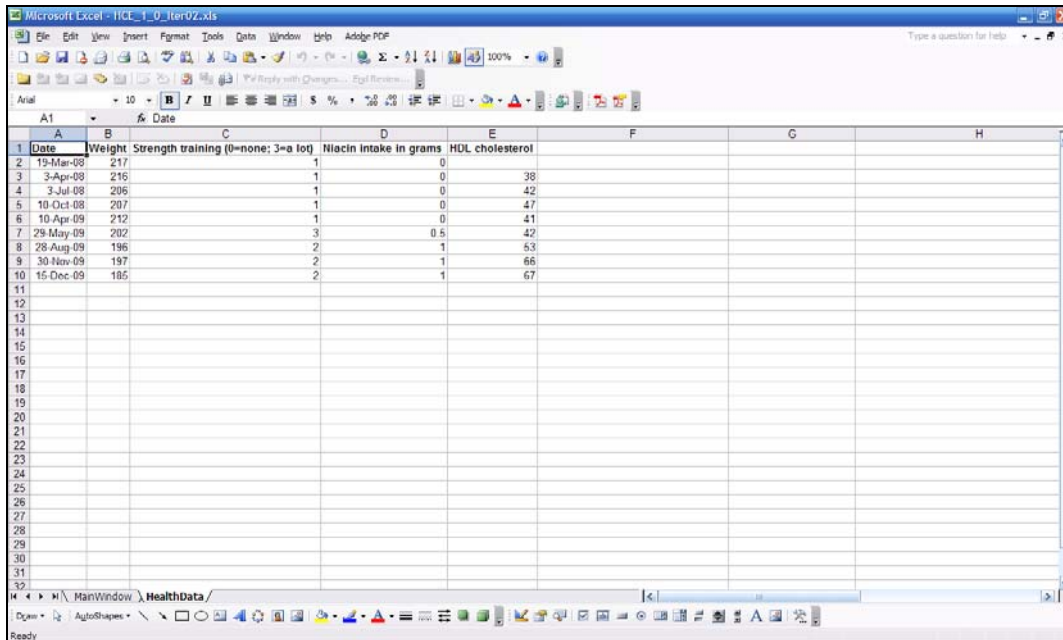**Figure 3.1. Multiple copies of HCE on the Windows desktop**

Typically a user will first download and run an initial copy of HCE. The user will then create a copy for analysis of, say, lipid panel data, including data on various variables (e.g., niacin intake, body weight, LDL and HDL cholesterol, triglycerides etc.). The user will then create a copy for analysis of exercise data, including data on various variables (e.g., repetitions and sets of various exercises, gains in strength etc.). Note that in this situation **the copies used for lipid panel and exercise data analyses will be kept separate**.

4. The HCE workbook and its sheets

HCE has two permanent sheets, which cannot be deleted, called MainWindow and HealthData. The MainWindow sheet contains the buttons that allow the user to access the software's features. The **HealthData sheet contains the data to be analyzed**. This is the health data that the user will enter. A sample dataset is provided with the trial version (see Figure 4.1). Users should simply enter their data in place of this sample data.

Figure 4.1. HealthData sheet with a sample dataset

Other sheets with coefficients of association and graphs will be generated by the software automatically. Prior to any analysis being conducted, only two sheets are available: the MainWindow and HealthData sheets. You can view the contents of any sheet by clicking on the respective sheet tab at the bottom. Each tab is identified by the name of the sheet it refers to.

5. The MainWindow sheet

The MainWindow sheet contains the buttons that allow the user to access the software's features (see Figure 5.1). The buttons under "Analyze" allow the user to access the main analysis features of the software. The buttons under "Delete" allow the user to delete certain sheets with analysis results, which can always be re-generated by the software. The buttons under "Licensing and help" allow the user to access web-based resources related to licensing and help on how to use to the software.

Figure 5.1. The MainWindow sheet

Clicking on **the button "Calculate associations and plot graphs"** generates several new sheets, showing them on two windows split vertically. The window on the left shows the Associations sheet, and the window on the right shows various Graphs( ) sheets. **The software assigns a number to each variable, showing it at the top of the Associations sheet, before each variable. This number appears within parentheses on the corresponding Graphs( ) sheet's name.** For example, the Graphs(3) sheet will show all the graphs associated with the variable numbered as 3, which are the graphs plotting that variable against all of the other variables in the HealthData sheet.

Clicking on **the button "Calculate associations only"** generates only the Associations sheet, which shows coefficients of association between each pair of variables. These coefficients of association are correlation coefficients, ranging from -1 to 1. This option does not split windows vertically. Clicking on **the button "Split associations/graphs windows vertically"** achieves that. If only the Associations sheet has been generated by the software (i.e., no graphs have been generated), the split windows view will show the Associations sheet on both the left and right windows.

6. The HealthData sheet

The HealthData sheet contains the data to be analyzed, which should be entered by the user. A sample dataset is provided with the trial version, illustrating how the health data should be entered. **The top of this sheet (i.e., the first row) contains the names of the health variables** (see Figure 6.1). **Below the names are the measurements** for each of the variables.

Figure 6.1. The HealthData sheet

The first variable on the left will typically be a variable that stores the dates at which the measurements were taken, but this is not a requirement. Only numeric measurements are allowed. The exception is for dates, which can be entered in any of the date formats handled by Excel. For example, one cannot enter as measurements the following: "low", "medium", or "high". However, one can enter the following: 1, 2 and 3. In this latter case, 1 may mean "low", 2 "medium", and 3 "high".

A minimum number of 5 measurements (i.e., rows) per variable are required. Within the scope of this constraint, **cells with missing values (e.g., the first cell below "HDL cholesterol") are allowed** and will not interfere with the calculation and graph plotting features of the software. **Missing values are not the same as zeroes.** A cell with a missing value corresponds to a variable for which a measurement was not taken at a particular point in time, while measurements for other variables were taken.

The **more measurements are available, the more reliable will be the coefficients of association** generated by the software. Generally, **30 measurements (i.e., rows) or more will lead to fairly reliable coefficients**. Still, **as little as 5 measurements per variable will already give the user an idea of how health variables are associated**. Similarly, measurement errors will be minimized by a large number of measurements.

Numeric variables can be continuous variables such as weight (measured on any unit of weight), or discrete variables. The latter (i.e., discrete variables) are normally measured on a Likert-type scale; examples are discrete scales going from 1 to 10, or from 0 to 3, or from -4 to 4. Typically discrete variables will be based on perceptions, such as one's perception of emotional stress or physical exertion. Continuous variables, on the other hand, will be objectively

measured; examples would be number of calories consumed, or the number of sets performed for a particular type of strength training exercise.

Perception-based variables are more likely to incorporate measurement error than objectively measured variables. This does not mean that objectively measured variables will always be free of measurement error. Sometimes error is built into an objective measurement approach where one variable is used as a "proxy" for another (e.g., weight gain, as a proxy for body fat gain). **On the other hand, perception-based variables are typically easier to measure** than objectively measured variables.

Some of the variables will be **independent variables**. These are health-related variables that are not affected by other variables in the dataset, and that likely affect one or more variables in the dataset. The affected variables may be mediating or dependent variables. For example, in a dataset containing diet-related variables, consumption of industrial vegetable oils (measured in, e.g., grams per day) would be an independent variable.
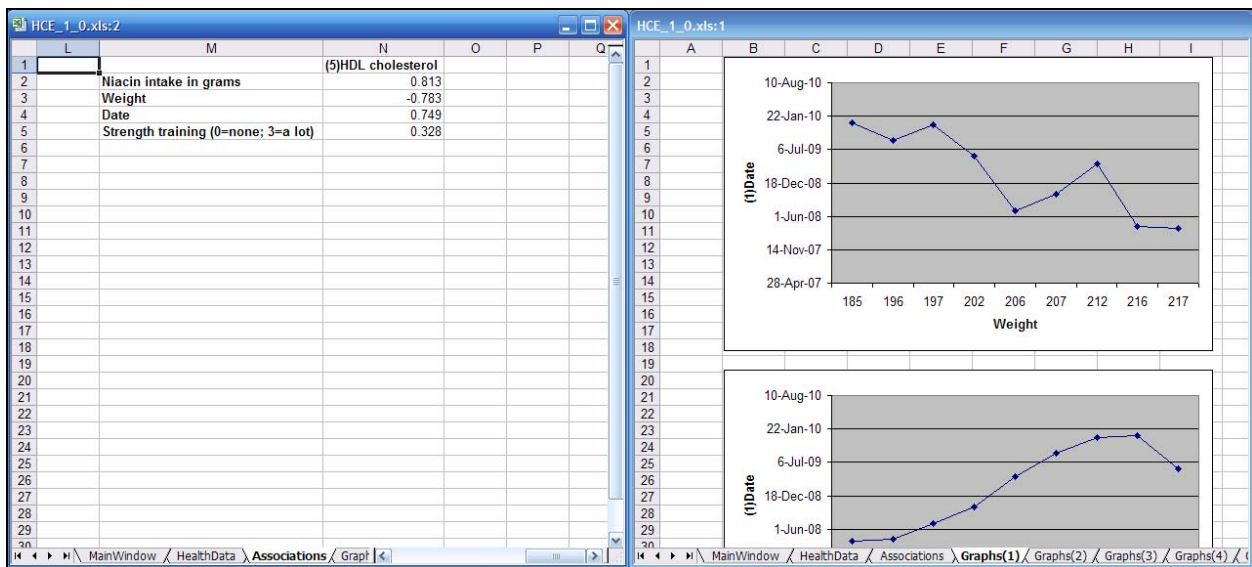
A **mediating variable** (a.k.a. "intervening variable") is a health-related variable that is affected by one or more variables in the dataset, and that also affects one ore more variables in the dataset. For example, in a dataset containing diet-related and lipid panel variables, body weight (measured in, e.g., lbs) would be a mediating variable. This is because body weight would likely be affected by consumption of industrial vegetable oils, and would likely affect HDL cholesterol.

A **dependent variable** is a health-related variable that is affected by one or more variables in the dataset, but that does not affect any other variable in the dataset. For example, in a dataset containing only standard lipid panel variables, total cholesterol (measured in, e.g., mg/dl) could be a dependent variable.

7. The Associations sheet

The Associations sheet is organized in areas that refer to "anchor" variables. The coefficients of association appear below each anchor variable in the Associations sheet (see Figure 7.1, where the anchor variable shown is HDL cholesterol). **The coefficients of association are correlation coefficients, which go from -1 to 1.** The closer they are to zero; the lower is the strength of the association between each variable on the left of the coefficient and the anchor variable. **The closer they are to -1 or 1, the greater is the strength of the association. Generally a coefficient of association should be of 0.6 or greater to suggest a strong association between a pair of variables.**
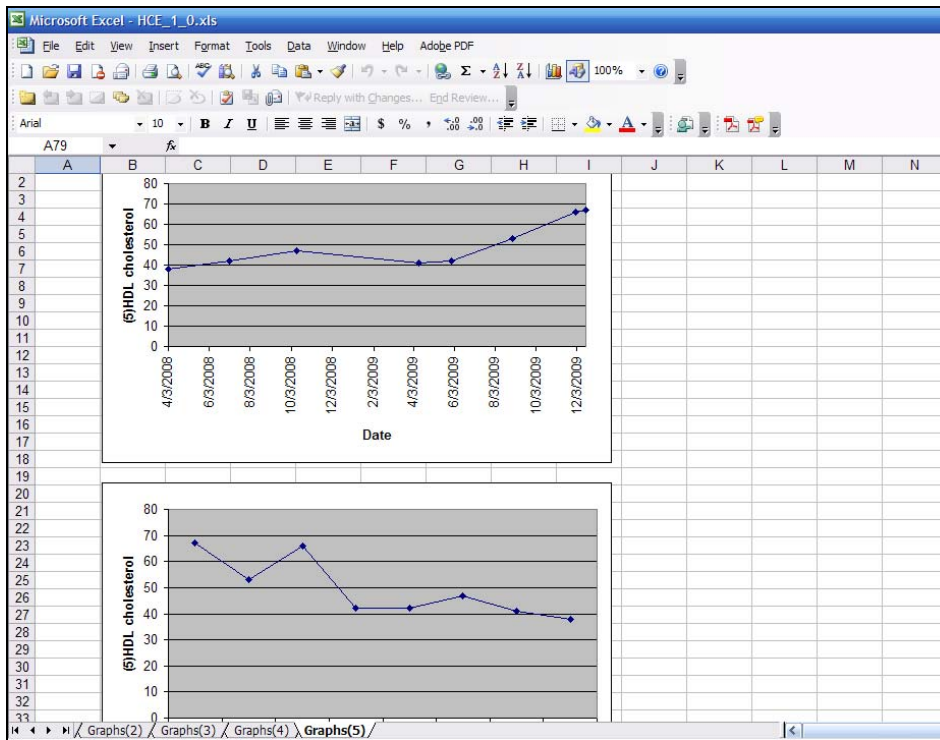
Figure 7.1. The Associations sheet

Negative coefficients of association refer to inverse relationships, where an increase in the variable to the left of the coefficient is associated with a decrease in the anchor variable, and vice versa. Coefficients of association are ordered based on strength, whether they are negative or positive.

Anchor variables always appear at the top of the Associations sheet, with a number preceding them. That number is also used in the Graphs( ) sheet that contains all the graphs associated with the anchor variable. **All variables in a dataset become anchor variables within the specific areas of the Associations sheet, with coefficients of association ordered below them.**

8. The Graphs( ) sheets

The Graphs( ) sheets can be viewed together with the Associations sheet, with windows split vertically and shown side by side, or by themselves as shown on Figure 8.1. **Each Graphs( ) sheet refers to one of the variables in the dataset. Those variables are numbered, with their respective numbers shown at the top of the Associations sheet.** Those are the numbers that appear within parentheses on the name of each Graphs( ) sheet. For example, all graphs for the variable numbered as 5 can be found in the Graphs(5) sheet.

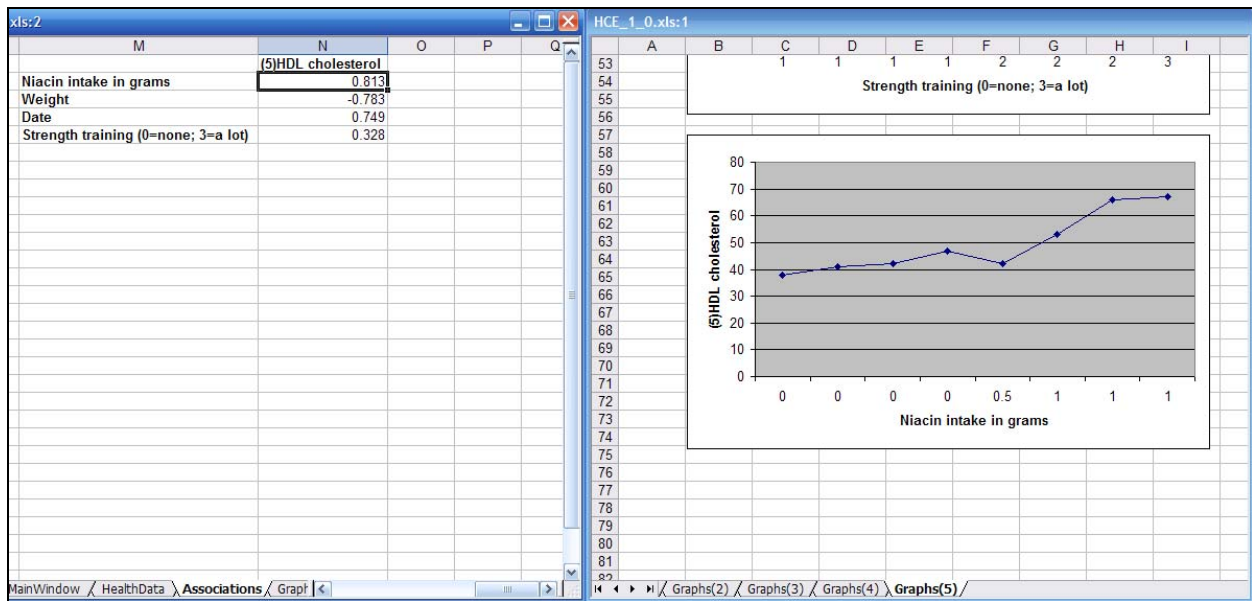**Figure 8.1. The Graphs( ) sheets**

Graphs complement coefficients of association in terms of the insights that they provide about relationships among variables. For example, a relationship between a pair of variables may be strong but nonlinear, with the shape of a "U curve", and yet the coefficient of association may be relatively low (i.e., below 0.6). The reason for this is that **the coefficients of association are correlation coefficients, which are inherently linear**. The coefficients of association are a measure of the extent to which a pair of variables vary together with reference to a regression line that best approximates the relationship.

9. Viewing associations and related graphs at the same time

Clicking on the button "Calculate associations and plot graphs" generates one Associations sheet and various Graphs( ) sheets, showing them on two windows split vertically as shown on Figure 9.1. **If the Associations and Graphs( ) sheets have already been generated, clicking on the button "Split associations/graphs windows vertically" will show them side by side.** Finally, users can split windows manually by using the appropriate Excel option; usually available under the menu options "Window" and "Arrange", with the appropriate setting being "Vertical".

Figure 9.1. Viewing associations and related graphs at the same time

Since graphs complement coefficients of association in the insights that they provide about relationships among variables, it is normally a good idea to view them side by side. While the coefficients of association express the extent to which two variables vary together, the graphs show the actual values. This is useful because **in some cases a very high coefficient of association will not translate into a noticeable effect in real life**. The telltale signs of this phenomenon, which is relatively rare, are points neatly distributed along a line with a very small slope.

To view side by side a coefficient of association for a pair of variables and the respective graph, do the following. First, locate the area for one of the variables in the Associations sheet, by looking for that variable as an anchor variable (listed at the top of the sheet). Second, go down the list until you find the coefficient of association that is on the right side of the other variable. Then go to the Graphs( ) sheet created for the anchor variable, and scroll down until you find the graph with the other variable listed below the horizontal axis.

Glossary

Anchor variable. A variable below which coefficients of association are organized in the Associations sheet. Anchor variables appear at the top of the Associations sheet, with a number preceding them. That number is also used in the Graphs( ) sheet that contains all the graphs associated with the anchor variable. All variables in a dataset become anchor variables within the specific areas of the Associations sheet organized below them.

Dependent variable. A health-related variable that is affected by one or more variables in the dataset, but that does not affect any other variable in the dataset. For example, in a dataset containing only standard lipid panel variables, total cholesterol (measured in, e.g., mg/dl) could be a dependent variable.

HCE. HealthCorrelator for Excel; i.e., the software to which this manual refers.

Independent variable. A health-related variable that is not affected by other variables in the dataset, and that likely affects one or more variables in the dataset. The affected variables may be mediating or dependent variables. For example, in a dataset containing diet-related variables, consumption of industrial vegetable oils (measured in, e.g., grams per day) would be an independent variable.

Mediating variable. Also known as "intervening variable". A health-related variable that is affected by one or more variables in the dataset, and that also affects one ore more variables in the dataset. For example, in a dataset containing diet-related and lipid panel variables, body weight (measured in, e.g., lbs) would be a mediating variable. This is because body weight would likely be affected by consumption of industrial vegetable oils, and would likely affect HDL cholesterol.