

4. Getting data in to PRIMER

- [Opening example data](#)
- [Importing data from Excel](#)
- [Post-import data checks](#)
- [Save your data & workspace](#)

Opening example data

Data from the Fal estuary

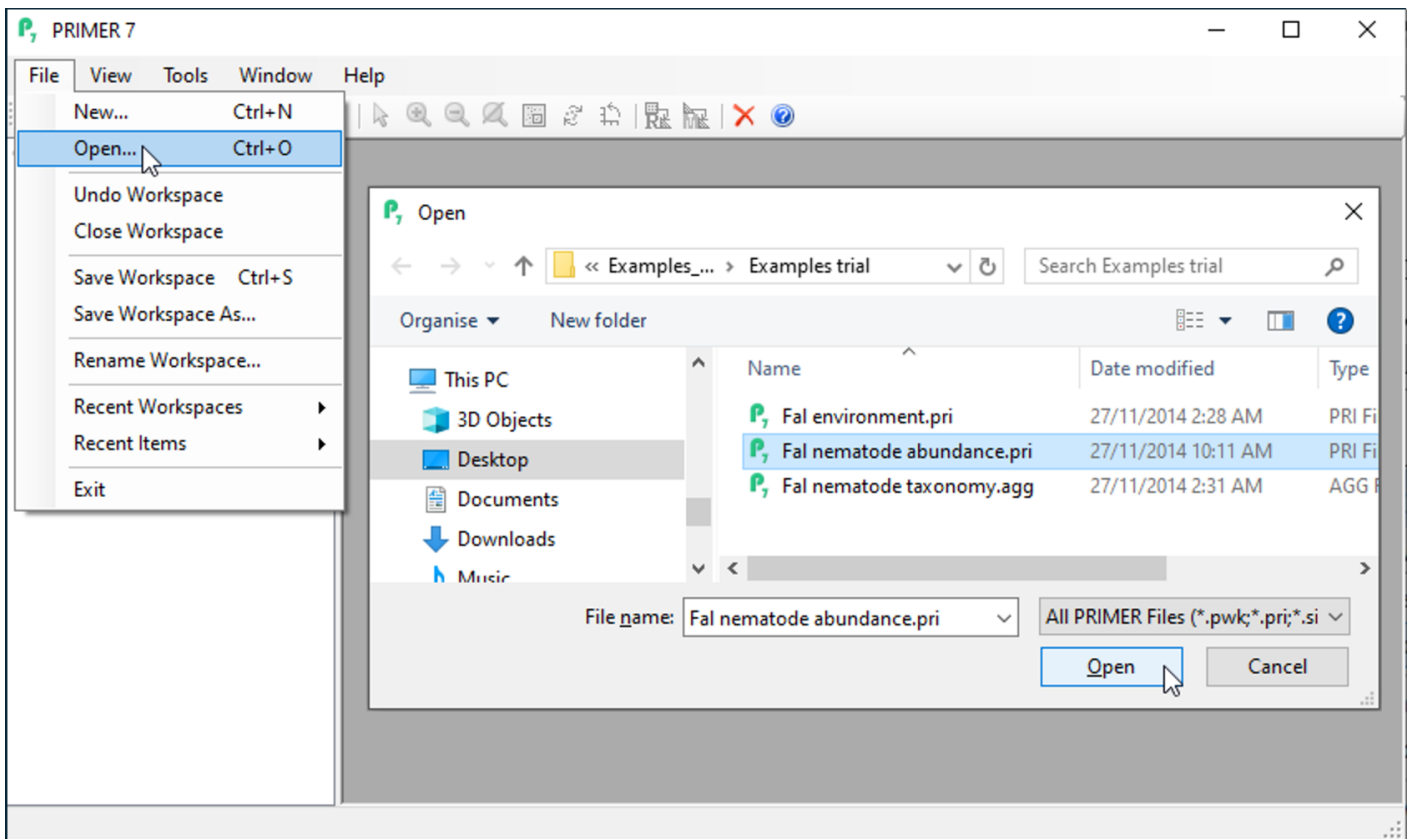
Example datasets in PRIMER can be obtained *via* the **Help** menu item. In trial mode, click **Help > Get Examples Trial...**, and you can download the following four files of example data (held in a folder called 'Examples Trial') to a location of your choice:

- **Fal environment.pri** (measured values for 12 environmental variables x 27 sites)
- **Fal environment.xls** (same as above, but in Excel format)
- **Fal nematode abundance.pri** (abundances of 62 nematode species x 27 sites)
- **Fal nematode taxonomy.agg** (taxonomic hierarchy for the 62 nematode species)

These data come from a study of benthic infaunal communities in soft sediments from 27 sites over five creeks of the Fal estuary, SW England ([Somerfield *et al.* \(1994a\)](#) , [Somerfield *et al.* \(1994b\)](#)). Sediments at these sites were contaminated to varying degrees by heavy metals, from historic mining activities. Both faunal counts (nematodes) and environmental measures were obtained from the same set of sites. (Note: the extension ***.pri** indicates a file containing a data matrix that has been saved in PRIMER 7's own internal (binary) format, unreadable by other software or earlier versions of PRIMER. Similarly, the ***.agg** extension indicates an aggregation-type file for PRIMER.)

Open the Fal data in PRIMER

To open the species-by-samples data matrix, launch PRIMER, then click **File > Open** from the main menu, navigate to the 'Examples trial' directory in the location you have specified, and select '**Fal nematode abundance.pri**'. Click **Open** to display the species matrix.



Alternatively, because this is a PRIMER file type (*.pri), you can instead use Windows Explorer to navigate to your specified folder and just double-click on the file name. This will launch a PRIMER session, with the data matrix open in the PRIMER desktop.

The screenshot shows the PRIMER 7 application window with the 'Fal nematode abundance' data matrix open. The matrix displays the abundance of various nematode species across seven samples (R1 to R7). The species are listed in the 'Variables' column, and the samples are listed in the 'Samples' column. The data is as follows:

Variables	R1	R2	R3	R4	R5	R6	R7
Anoplostoma vivip	0	0	0	0	0	0	0
Halalaimus gradis	0	0	0	0	0	0	0
Halalaimus longica	0	0	0	0	0	0	0
Oxystomina elonga	0	0	0	0	0	0	0
Viscosia viscosa	0	0	0	0	0	0	0
Tripyloides gradis	149	181	385	289	170	614	
Atrochromadora mi	0	0	0	0	0	0	0
Chromadora macro	0	4	29	63	263	123	
Chromadora nudica	0	0	0	0	0	0	7
Chromadorella ?du	0	0	0	0	0	0	0
Chromadorita nana	0	0	0	0	0	0	0
Chromadorita tenta	0	0	0	0	0	0	0
Dichromadora geop	5	0	0	0	0	0	7
Hypodontolaimus h	40	44	25	18	5	14	

Properties of the data

Click on **Edit > Properties** and you will see that PRIMER-format *.pri sheets carry other information about the data matrix as well, including:

- Title,
- Data type,
- Array size (number of columns and rows),
- Orientation (whether samples are found in columns or rows),
- a Description, and
- a History of what pre-treatments (such as a transformation) have been applied to them.

Sample Data Properties

Title:
Fal estuary nematodes

Data type:
☒ Abundance
☐ Biomass
☐ Environmental
☐ Unknown/other

History:

Samples as:
☒ Columns
☐ Rows

Number of columns: 27
Number of rows: 62

Description:
Abundance of nematodes (counts from identifications in a random sub-sample multiplied up by total nematode count for each sample, so has the statistical properties of density rather than pure count).
Data from meiofauna cores of the sediments of 27 sites over 5 creeks of the Fal estuary, Cornwall, UK.

OK Cancel Help

Factors associated with the data

Click on **Edit > Factors**, and you can see that a subsidiary sheet of three factors is also linked to this worksheet: 'Creek', a single-letter abbreviation for the creeks, the full 'Creek name' and a numeric 'Position' factor identifying the location of each sampling site down each creek.

Factors

Edit Fill

Add... Combine... Rename... Reorder... Delete... Key... Import... OK Cancel Help

Label	Creek	Creek name	Position
R1	R	Restronguet	1
R2	R	Restronguet	2
R3	R	Restronguet	3
R4	R	Restronguet	4
R5	R	Restronguet	5
R6	R	Restronguet	6
R7	R	Restronguet	7
M1	M	Mylor	1
M2	M	Mylor	2
M3	M	Mylor	3
M4	M	Mylor	4
M5	M	Mylor	5
P1	P	Pill	1
P2	P	Pill	2
P3	P	Pill	3
P4	P	Pill	4
P5	P	Pill	5
J1	J	St Just	1
J2	J	St Just	2

Note that additional factors could be added here by clicking **Add**, and also combinations of levels of existing factors can be created by clicking on **Combine**.

Importing data from Excel

Step 1. Ensure your data are in a format suitable for import into PRIMER

Suppose we have a dataset in Excel that is already in a suitable format for import into PRIMER. The environmental data from the Fal estuary provides an example of this. These data are found in the file 'Fal environment.xls' and consist of values for each of 12 environmental variables measured from sediments collected from 27 sites across 5 tidal creeks in the Fal Estuary (available from within PRIMER by clicking **Help > Get Examples Trial...**, as seen in the last section).

Fal estuary environmental variables											
	R1	R2	R3	R4	R5	R6	R7	M1	M2	M3	M4
% silt/clay	87	76	71	73	45	70	61	97	92	93	
% organic carbon	6.8	6.5	5.7	5.7	3.9	8.7	6.5	8.5	9	8.5	
Ag	4.94	4.91	3.27	3.04	3.09	3.43	2.73	2.34	2.2	2.48	
Cd	3.75	4.45	2.49	2.61	1.44	2.59	1.91	1.57	1.48	1.46	
Co	27.4	27.2	21.5	18.7	17.3	22.2	18.8	13.3	12.8	12.2	
Cr	45.4	43.8	37.9	32.4	34	41.3	36.4	66.3	56.5	63	
Cu	3302	3373	2387	2176	2040	2452	1997	1354	1246	1274	
Fe	67862	68098	53118	49505	47250	55256	49826	42261	41326	41584	
Mn	611	586	521	521	507	539	488	405	411	392	
Ni	37.3	35.9	28.1	26.5	27.6	29.2	25.3	33.8	31.7	32.6	
Pb	271	265	192	180	165	206	185	196	179	186	
Zn	5279	5873	3583	3328	2252	3475	2900	1460	1357	1481	
Creek	R	R	R	R	R	R	R	M	M	M	M
Creek name	Restrongu	Restrongu	Restrongu	Restrongu	Restrongu	Restrongu	Restrongu	Mylor	Mylor	Mylor	Myl
Position	1	2	3	4	5	6	7	1	2	3	

Important things to note about this file:

- There is a *title* for the dataset ('Fal estuary environmental variables') in the very first (upper left-hand) cell (A1). This title is optional, but handy as a naming convention.
- The cell immediately under the title (cell A2) is *empty*.
- There are *column labels* ('R1', 'R2', ...) in row 2. These are unique labels for the sampling units (Sites in this case).

- There are *row labels* ('%silt/clay', '% organic carbon', ...) in column A. These are unique names associated with each variable.
- The entries for every cell in the matrix of data itself (beginning with cell B3) all contain *numerical values only*. There are no non-numeric characters. This means that you may not use 'NaN' or 'NAN' to denote missing values. If data are missing from a cell, then it should be left **blank**. In addition, symbols such as '<' or '>' (for 'less than' or 'greater than') are similarly not permitted or accepted as valid data values within the data matrix.
- In this example, the variables are rows and the sampling units (sites) are columns. It is perfectly ok to have this formatted the other way around, with variables as columns and sampling units as rows. You will specify the orientation of your data matrix explicitly when you import your Excel file into PRIMER.

This format must be adhered to precisely, with no extra blank rows or columns, or extra headers, otherwise PRIMER will not be able to open it successfully.

Other things to note:

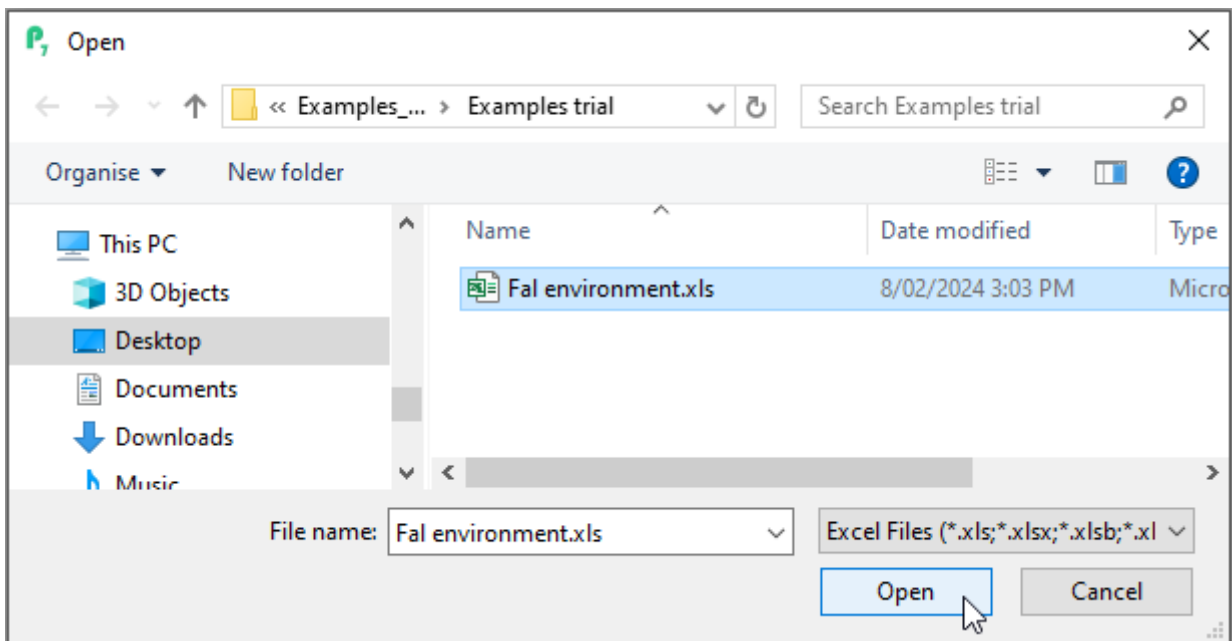
- **Factors:** You can label the sampling units as belonging to a level of one (or more) factors by **skipping a line** at the bottom of the matrix and placing this 'factor information' there. In the above image you can see that there are three factors: 'Creek' (row 16), 'Creek name' (row 17) and 'Position' (row 18).
- **Indicators:** You can similarly label variables as belonging to particular groups in the same manner; this is done along the *other* margin of the data matrix (e.g., after skipping a column, for this example). This might be useful for doing analyses on subsets of variables belonging to different types, such as physical vs chemical variables. In a case where variables are species, one might want to consider subsets of variables corresponding to families, functional groups, etc.

Inclusion of one (or more) *factors* (to specify groups of samples) or *indicators* (to specify groups of variables) is optional. If you have more than one factor, then these are given one after the next (in adjacent rows); do not put blank rows between multiple factors. The initial single blank row (or column) is there simply to demarcate the difference between the data matrix itself and additional information about the data matrix upon import.

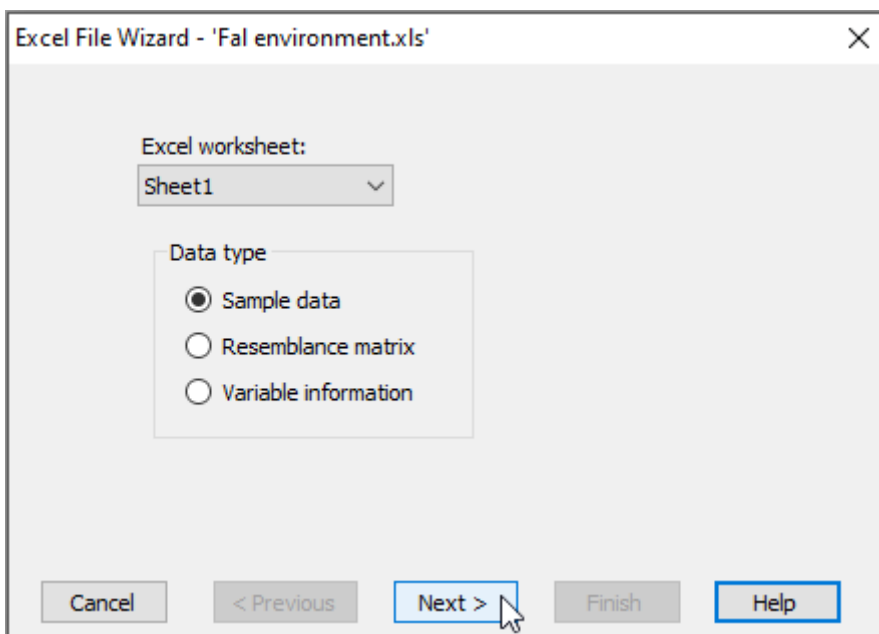
Step 2. Open PRIMER and import the data from Excel

Once your Excel file is ready, open up PRIMER and choose **File > Open**. Look at the bottom of the dialog box and you will see next to the words 'File name:' that the only files that PRIMER can see is: 'All PRIMER Files...'. Click on 'All PRIMER Files...' and change this to 'Excel Files...'. Once you have done this, you should be able to browse and see the Excel data file that you want.

1. Click on the name of your Excel file in the browser (here it is 'Fal environment.xls'), then click **Open**.



- This will initiate PRIMER's Excel File data-import Wizard. Choose the name of the specific sheet within your Excel file that contains your data and the type of data you are importing. Here, we have (Excel worksheet: **Sheet1**) & (Data type **Sample data**), then click **Next >**.



- Choose the correct orientation, type of data and the meaning of blank entries (if any). For this example, we have (Orientation **Samples are columns**) & (Data type **Environmental**) & (Blank = **Missing value**), then click **Finish**.

Excel File Wizard - 'Fal environment.xls'

☒ Title
☒ Row labels

Orientation
☒ Samples as columns
☐ Samples as rows

Data type
☐ Abundance
☐ Biomass
☒ Environmental
☐ Unknown/other

Blank =
☒ Missing value
☐ Zero

Cancel < Previous Next > Finish Help

4. You will now see your data file has been imported and is nicely displayed in the PRIMER workspace. It appears in its own window, and its name also appears in the 'Explorer tree'-type window shown on the left-hand side of the PRIMER desktop.

PRIMER 7

File Edit Select View Wizards Pre-treatment Analyse Plots PERMANOVA+ Tools Window Help

Workspace
Fal environment

Fal environment
Fal estuary environmental variables
Environmental

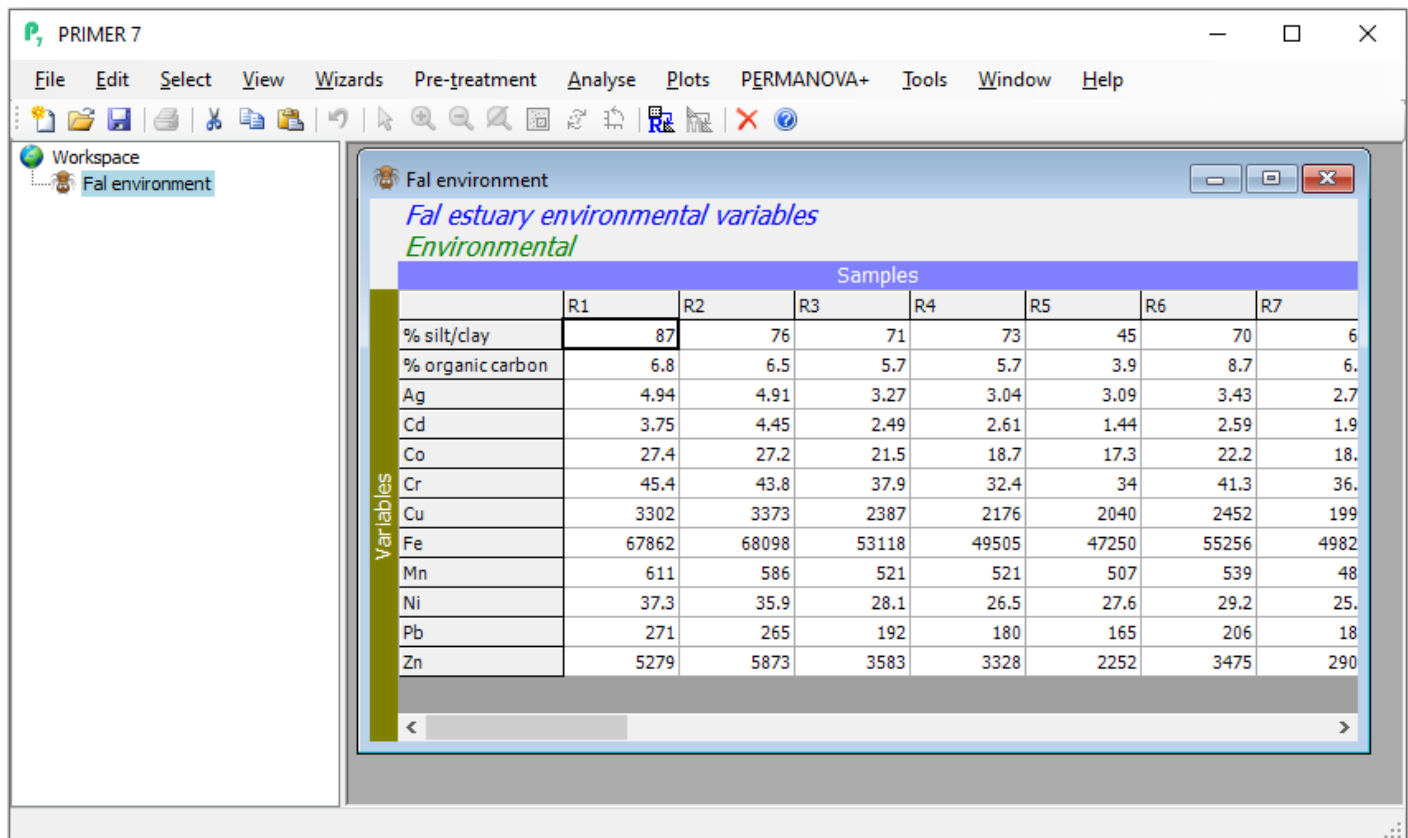
	Samples						
	R1	R2	R3	R4	R5	R6	R7
% silt/clay	87	76	71	73	45	70	6
% organic carbon	6.8	6.5	5.7	5.7	3.9	8.7	6.
Ag	4.94	4.91	3.27	3.04	3.09	3.43	2.7
Cd	3.75	4.45	2.49	2.61	1.44	2.59	1.9
Co	27.4	27.2	21.5	18.7	17.3	22.2	18.
Cr	45.4	43.8	37.9	32.4	34	41.3	36.
Cu	3302	3373	2387	2176	2040	2452	199
Fe	67862	68098	53118	49505	47250	55256	4982
Mn	611	586	521	521	507	539	48
Ni	37.3	35.9	28.1	26.5	27.6	29.2	25.
Pb	271	265	192	180	165	206	18
Zn	5279	5873	3583	3328	2252	3475	290

Now that the data are in PRIMER, it is a good idea to do some post-import data checks.

Post-import data checks

Check the orientation

After import, make sure you have specified the orientation correctly by examining the labels on the columns and rows of the data frame. For example, after importing the Fal environmental data from Excel (see the previous page), you can see that the columns are 'Samples' (a periwinkle-coloured strip across the top) and rows are 'Variables' (an olive green-coloured strip along the left margin).



PRIMER 7

File Edit Select View Wizards Pre-treatment Analyse Plots PERMANOVA+ Tools Window Help

Workspace
Fal environment

Fal environment

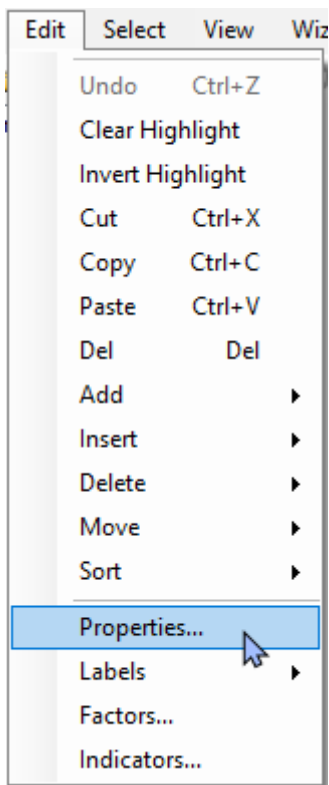
Fal estuary environmental variables
Environmental

	Samples						
	R1	R2	R3	R4	R5	R6	R7
% silt/clay	87	76	71	73	45	70	6
% organic carbon	6.8	6.5	5.7	5.7	3.9	8.7	6.
Ag	4.94	4.91	3.27	3.04	3.09	3.43	2.7
Cd	3.75	4.45	2.49	2.61	1.44	2.59	1.9
Co	27.4	27.2	21.5	18.7	17.3	22.2	18.
Cr	45.4	43.8	37.9	32.4	34	41.3	36.
Cu	3302	3373	2387	2176	2040	2452	199
Fe	67862	68098	53118	49505	47250	55256	4982
Mn	611	586	521	521	507	539	48
Ni	37.3	35.9	28.1	26.5	27.6	29.2	25.
Pb	271	265	192	180	165	206	18
Zn	5279	5873	3583	3328	2252	3475	290

If you happen to get this the wrong way around (e.g., if your variables are actually columns instead of rows), this can easily be changed (swapped around) by choosing **Edit > Properties** and toggling the radio button for 'Samples as' to either '\$\bullet\$Columns' or '\$\bullet\$Rows', whichever is appropriate.

Check the properties, factors and indicators

To be sure that the import has been fully successful, including all data points, factors and indicators that may have been included in your original Excel file, you can see additional information attached to your data matrix by clicking on your imported dataset in PRIMER, and doing the following:



- Look at the data properties, size of the matrix, etc.: Click **Edit > Properties...** Note that you can add a useful 'Description' of your data into this dialog if you like. (For the Fal environmental dataset, we can see there are 12 variables and 27 sites, etc.).

Sample Data Properties

Title:
Fal estuary environmental variables

Data type

☐ Abundance
☐ Biomass
☒ Environmental
☐ Unknown/other

History:

Samples as

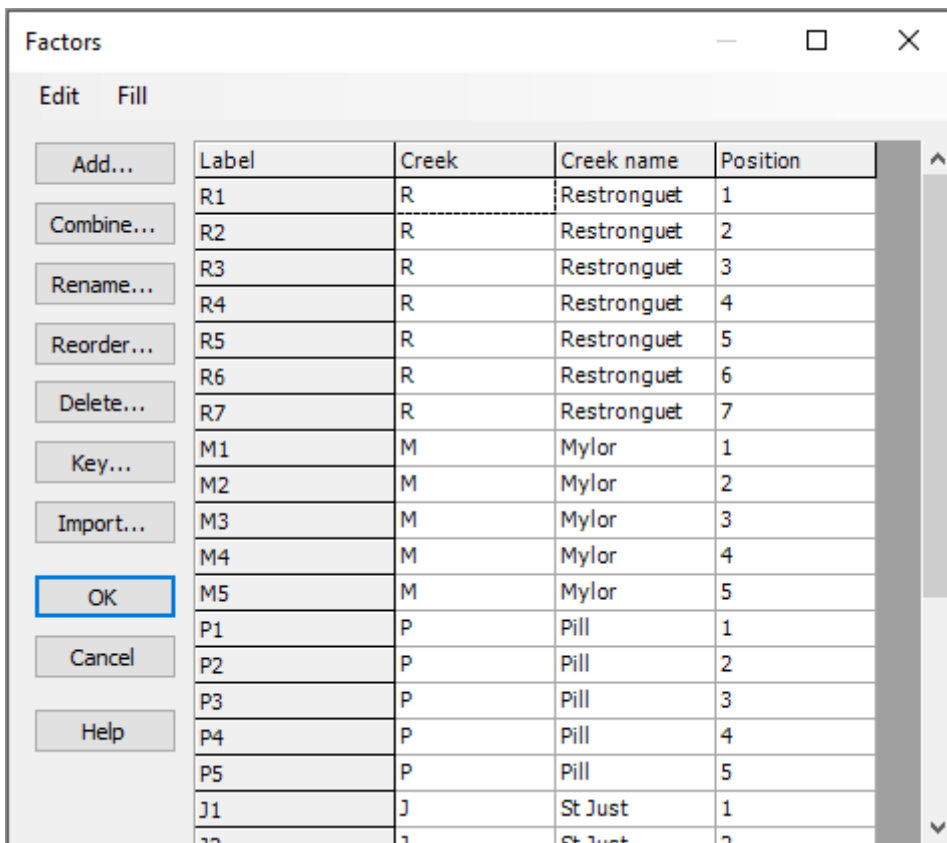
☒ Columns
☐ Rows

Number of columns: 27
 Number of rows: 12

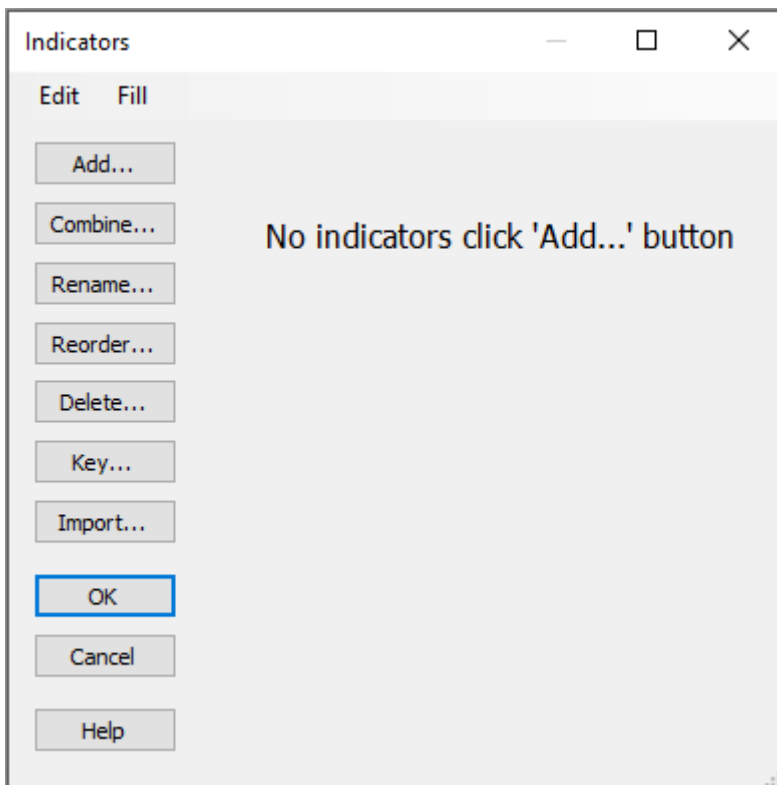
Description:

OK Cancel Help

- Look at the Factors (if any): Click **Edit > Factors...** (For the Fal environmental dataset, you will see the same three factors of 'Creek', 'Creek name' and 'Position' that we saw in the Excel file).



- Look at the Indicators (if any): Click **Edit** > **Indicators....** (For the Fal environmental dataset, there were no indicators, but you could add some here, if you wish).

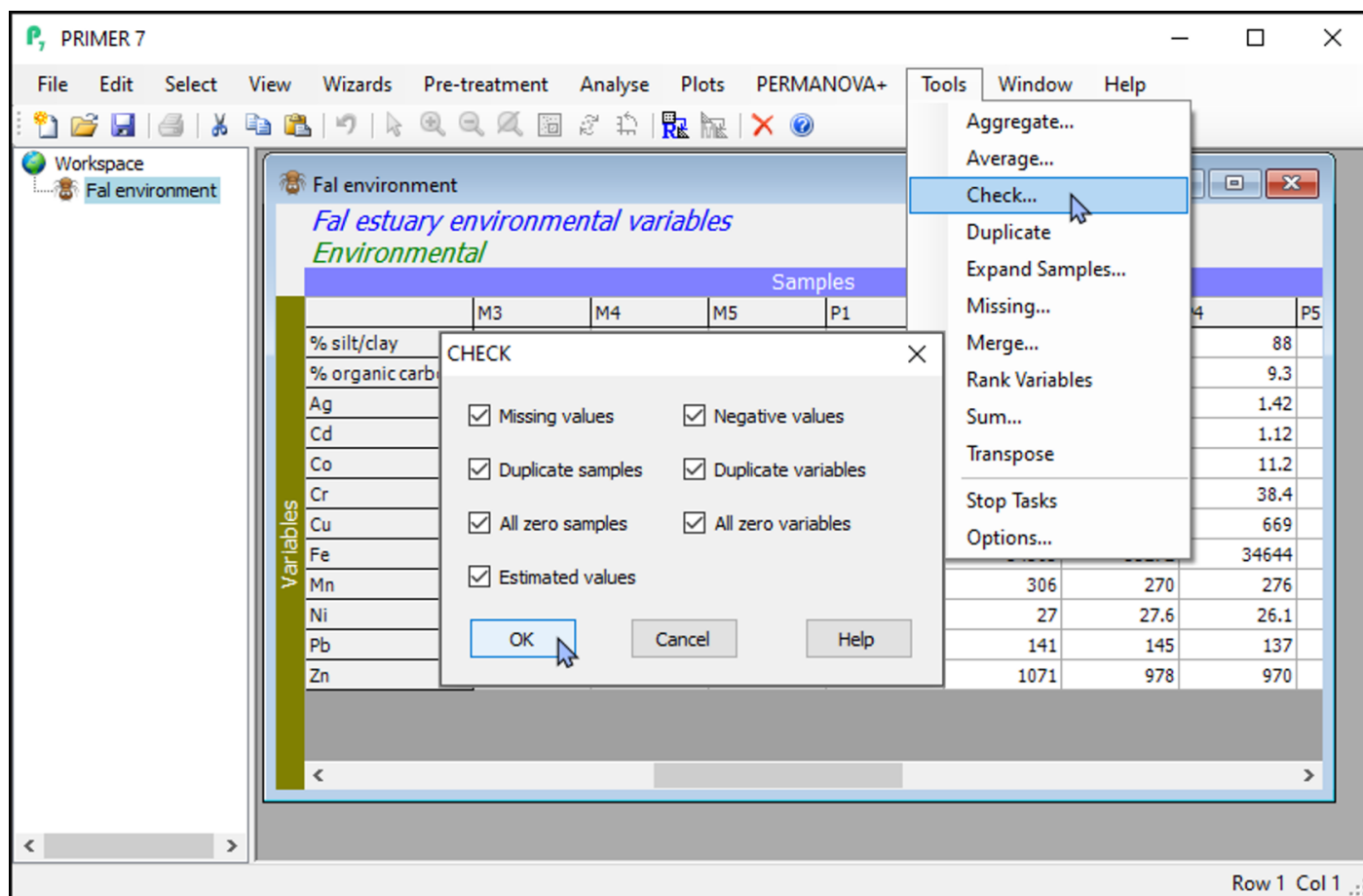


Run PRIMER's data-checking tool

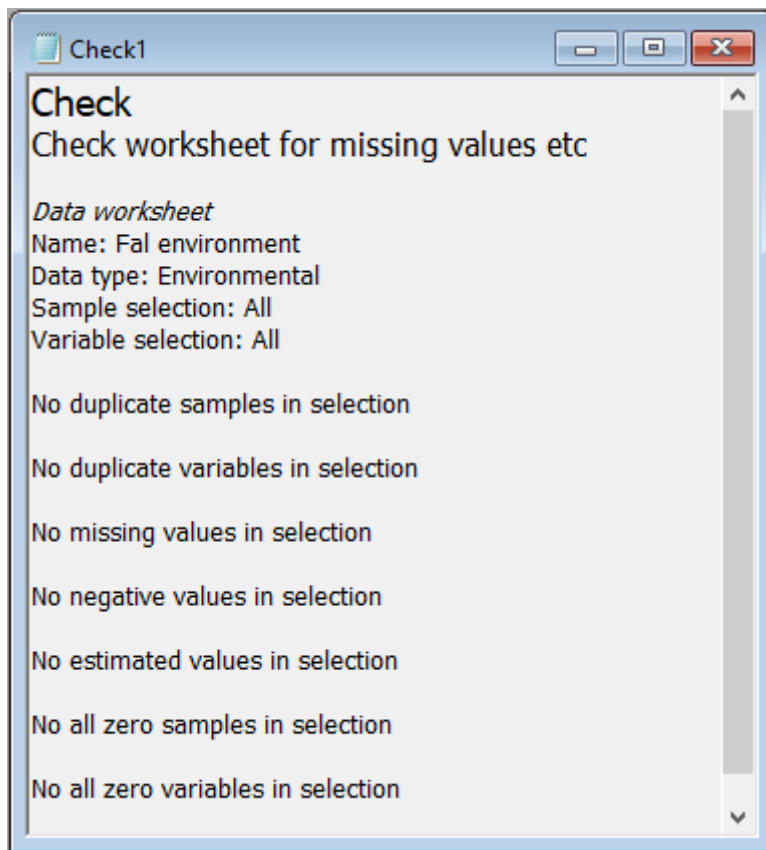
As an additional option, you can run PRIMER's internal data-checking tool to find and identify certain other features that might be present in your data, including:

- Missing values,
- Negative values,
- Duplicate samples,
- Duplicate variables,
- All-zero samples,
- All-zero variables, and/or
- Estimated values

To run this routine, start by clicking on your datasheet, then click on **Tools > Check...**:



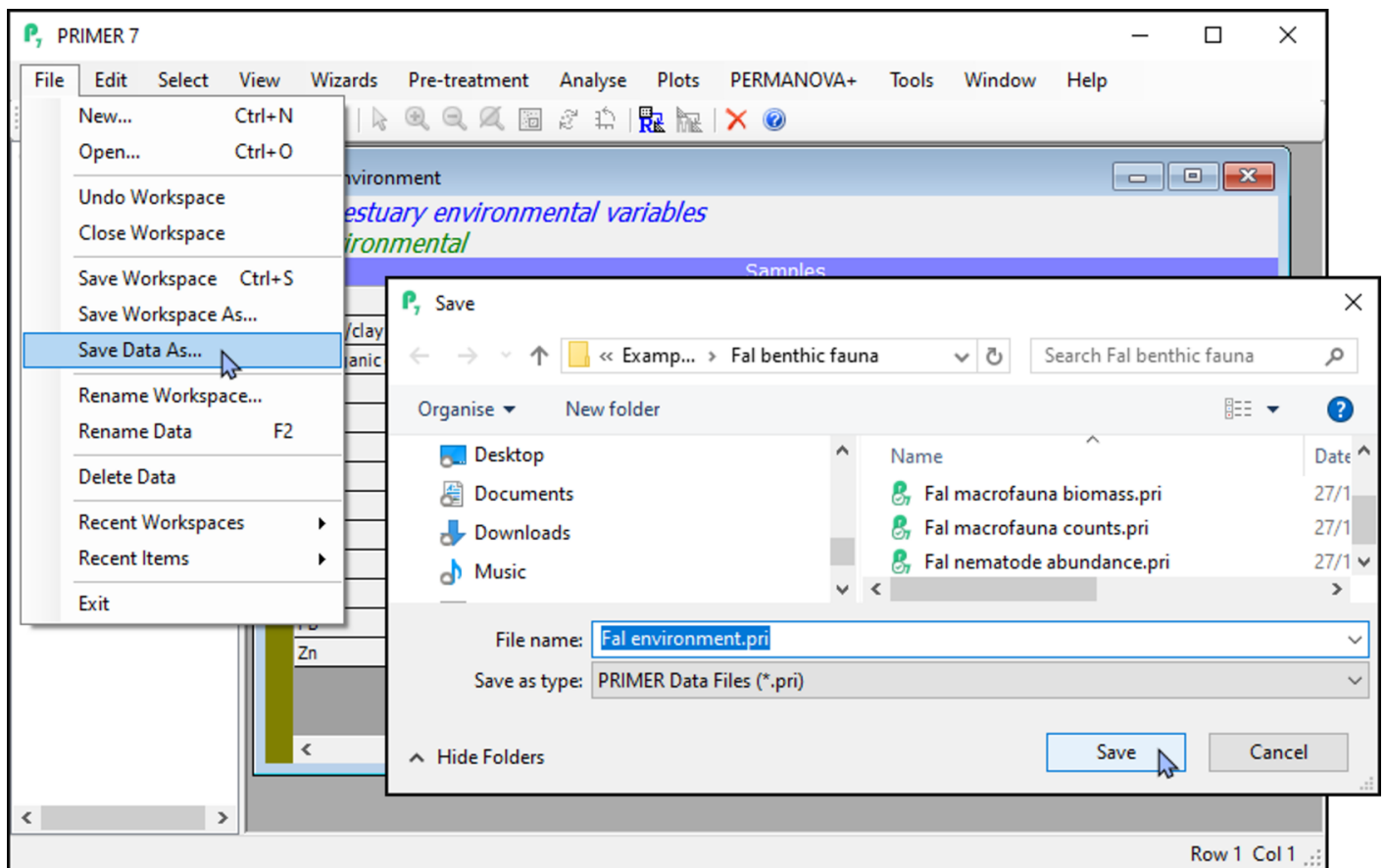
For the Fal environmental data, none of these features occurred (see below), and we are ready to proceed with subsequent analyses.



Save your data & workspace

Save your data in PRIMER (*.pri) format

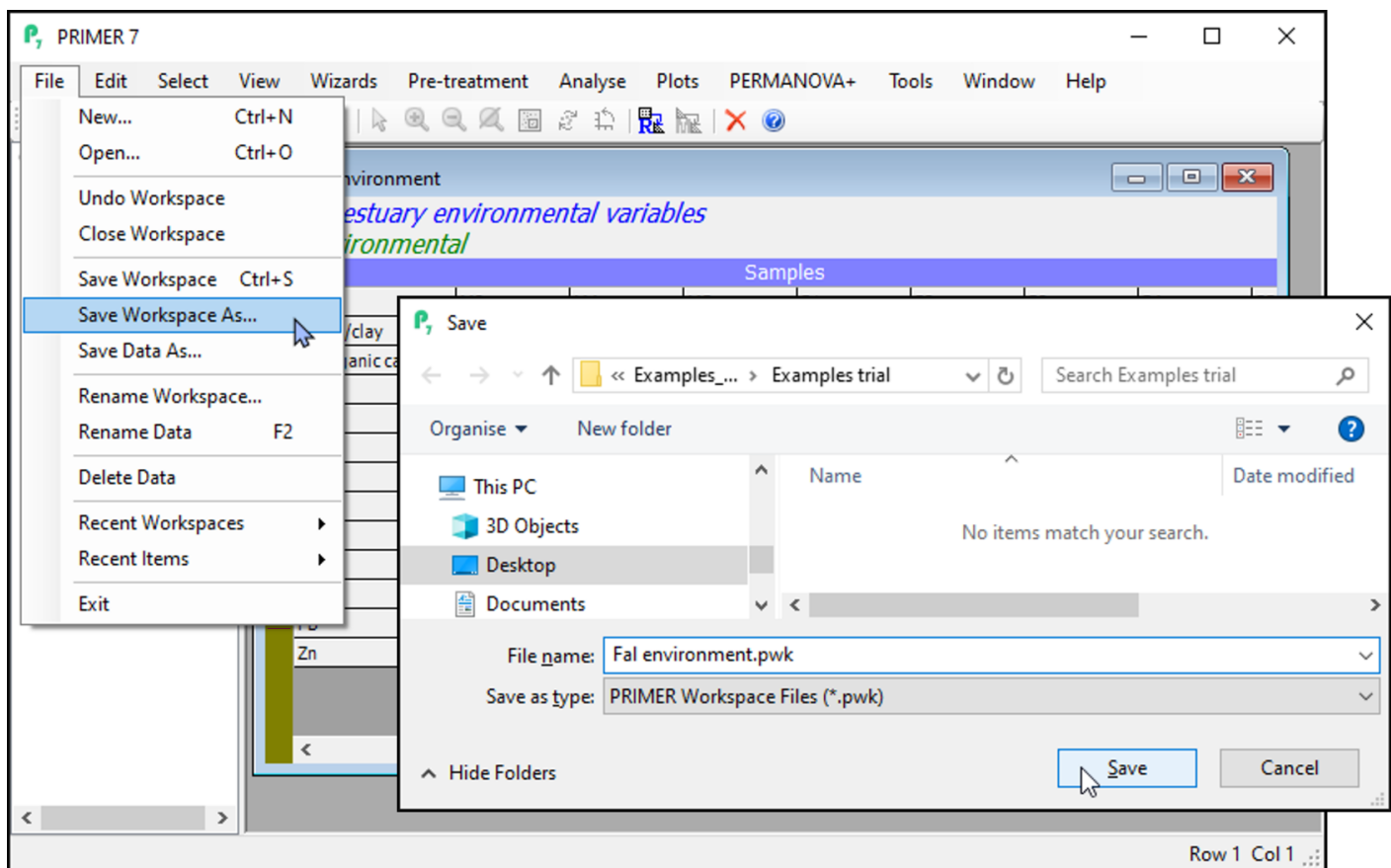
To save a data sheet in PRIMER (*.pri) format, click on the data sheet inside the PRIMER workspace you want to save and click **File > Save Data As...**



Note the default file-type for saving a data file is 'PRIMER Data Files (*.pri)', as shown in the box labeled 'Save as type:' in the 'Save' dialog. You can change this (by clicking in the 'Save as type:' box) in order to save the data in some other format of choice, such as *.txt, *.csv *.xls, *.xlsx, or in a PRIMER format compatible with older versions of the PRIMER software.

Save your entire PRIMER workspace (*.pwk)

Generally, you will want to save your data *plus* all of the work you have done in PRIMER to analyse that dataset. To save the entire PRIMER workspace, click **File > Save Workspace As...**



Type the file name you wish (e.g., **Fal environment.pwk**), and click **Save**. This will save the *entire* PRIMER workspace (not just the datasheet), including all elements you may have created during the PRIMER session (e.g., data, graphics, analyses, etc.), all of which you are able to navigate and see listed (in hierarchical fashion) within the Explorer tree window on the left-hand side of the open PRIMER workspace.

A PRIMER workspace file is identifiable by using the file extension ***.pwk** in the file name. Double-clicking on a file with this extension will open up that entire PRIMER workspace within the PRIMER software.