

1.5 Environmental data: summary stats

For environmental data, we might choose to calculate different sorts of summary statistics than the kinds of things we would want to know about biotic data consisting of counts. For count data, quantities like the numbers of zeros, singletons, doubletons and frequencies of occurrence might well be of interest. However, these are not typically meaningful for variables that have been recorded as (effectively) continuous values, such as temperature, dissolved oxygen, etc. Instead, when we deal with environmental data (or, more generally, continuous quantitative variables), it may be helpful to know what the smallest non-zero value is, or to know the degree of skewness. These can aid in determining an appropriate transformation (e.g., to obtain approximate symmetry).

Let's look at some environmental data from a study of benthic soft-sediment assemblages in the Firth of Clyde, SW Scotland ([Pearson & Blackstock \(1984\)](#)). The abundance and biomass of 84 macrofaunal species as well as contaminant data (organic enrichment and concentrations of heavy metals in the sediment) were sampled at a series of 12 sites along a transect that passed through the sewage-sludge disposal ground at Garroch Head. Open the data file named 'Clyde_environment.pri' in PRIMER (found inside the 'Examples_P8 > Clyde_macrofauna' folder).

PRIMER

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

Workspace

Clyde_environment

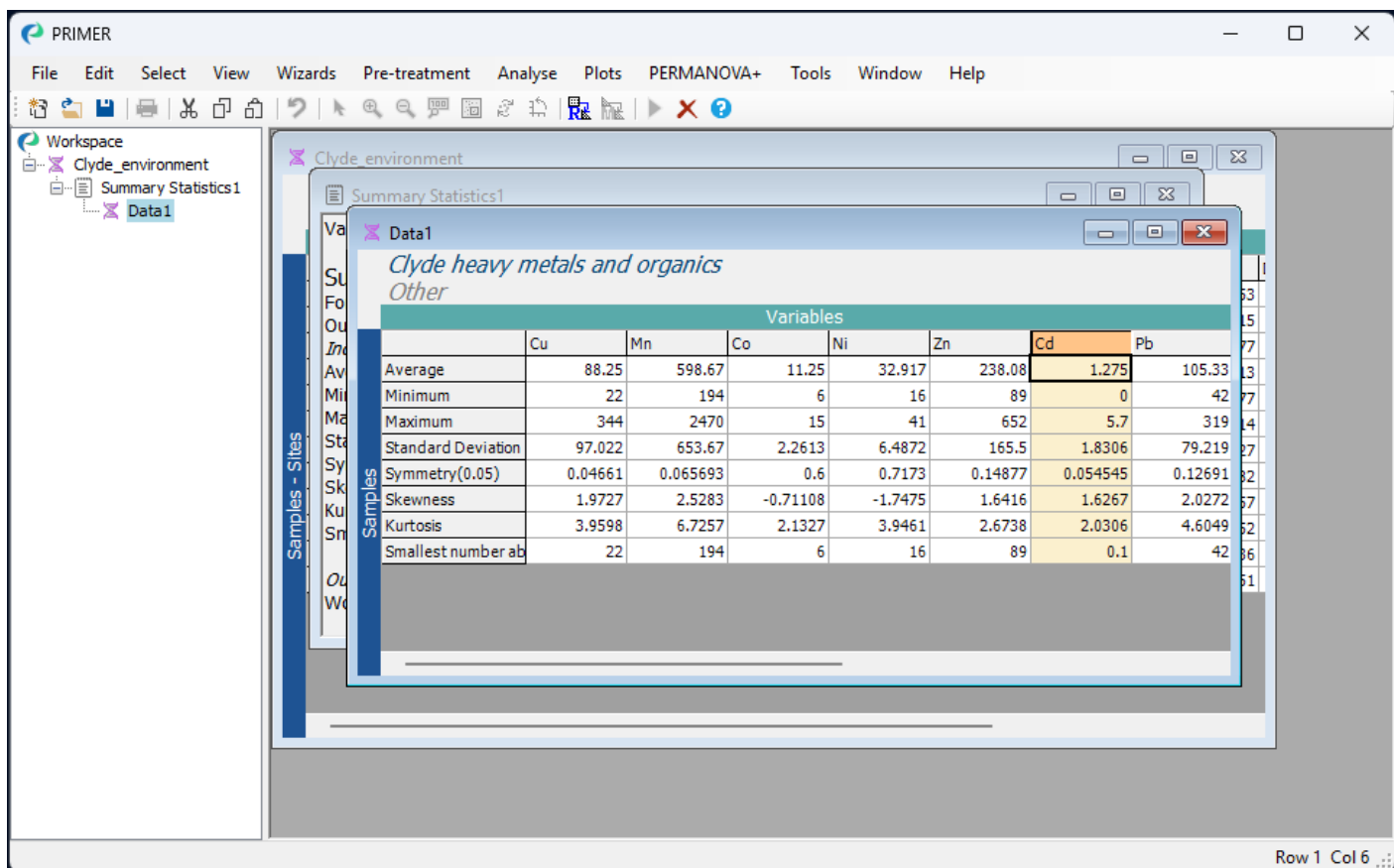
Clyde_environment

Clyde heavy metals and organics
Environmental

	Variables								
	Cu	Mn	Co	Ni	Zn	Cd	Pb	Cr	
S1	26	2470	14	34	160	0	70	53	
S2	30	1170	15	32	156	0.2	59	15	
S3	37	394	12	38	182	0.2	81	77	
S4	74	349	12	41	227	0.5	97	113	
S5	115	317	10	37	329	2.2	137	177	
S6	344	221	10	37	652	5.7	319	314	
S7	194	257	11	34	425	3.7	175	227	
S8	127	246	10	33	292	2.2	130	182	
S9	36	194	6	16	89	0.4	42	57	
S10	30	326	11	26	108	0.1	44	52	
S11	24	439	12	34	119	0.1	58	36	
S12	22	801	12	33	118	0	52	51	

Of course, tools such as histograms and draftsman plots (found under the **Plots** menu) are very useful for visualising the distributions of values for the individual variables. Summary statistics complement these visual tools, yielding some important additional detailed information. Click **Tools > Summary Stats...**, then choose to output the following statistics (shown in the dialog below): average, minimum, maximum, standard deviation, symmetry (0.05), skewness, kurtosis, and smallest number above a threshold of 0.

The resulting output file (called 'Data1' and shown below) indicates that many of the variables are right-skewed, with positive values for skewness, and values less than 0.5 for the symmetry statistic, while others (Co, Ni) are apparently left-skewed (negative skewness and symmetry > 0.5). Depth, however (Dep) is a fairly flat (platykurtic) variable (with negative excess kurtosis). Also evident is the fact that values of cadmium concentration (Cd) reach a minimum value of zero, and that the smallest value recorded above zero is 0.1. Thus, if we were inclined to transform the values for Cd (e.g., to make its distribution more symmetric), we might consider a log transformation such as $\log(y_i + c)$, where $c = 0.1$.



(Note: the variable 'Cd' has been highlighted in the above image of 'Data1' by clicking on the heading for that column. Highlighting variables (or samples), can be very useful for selecting subsets of data, or for applying a transformation to a subset of variables.)

Revision #38

Created 17 November 2024 21:57:15 by Marti

Updated 1 December 2025 21:09:02 by Marti