

## UNISTAT: STATISTICAL AND GRAPHICS PACKAGE FOR GEOCHEMISTS

Larry Nolan  
Geochemistry and Geophysics Section

### ABSTRACT

*UNISTAT is an interactive computer program that calculates descriptive statistics of geochemical data. The program creates graphical presentations of histograms and cumulative plots, for visual interpretation of dataset characteristics. Symetry, population distributions, range and outliers are readily interpreted from the histogram, and the cumulative frequency plot enables element concentration breaks to be uniquely identified. Program output is to the computer graphics display screen, selected printers and pen plotters.*

### INTRODUCTION

UNISTAT (UNIvariate STATistics) is an interactive computer program for the IBM-PC and compatible microcomputers. The program is designed to assist in the processing, analyzing and interpretation of geochemical data on an individual element basis (univariate), by providing basic dataset statistics combined with publication-quality graphics output. Descriptive statistics (minimum, maximum, mean, median, standard deviation and coefficient of variation) are computed using selected options of upper-lower data cutoff limits and logarithmic transformation, to produce histograms and cumulative frequency plots on dot matrix printers and pen plotters.

The histogram provides immediate information on the frequency distribution of an element. It determines if the element frequency distribution is symmetrical (normal) or skewed, unimodal (one population) or polymodal (several populations) and whether outliers are present, and also the range of the data. The histogram can be constrained by selecting a data subset using the minimum and maximum cutoff limits to remove outliers, and also by using actual values or their logarithms (to base 10). Also, the number of class intervals may be selected to optimize the histogram, to show the data as generalized large intervals or smaller more detailed groups of the data. A default number of class intervals is calculated based upon an interval of one quarter of the standard deviation.

The cumulative frequency plot provides a graphics presentation of the cumulative frequency distribution (cfd) on a normal probability scale, and the use of this plot has been discussed by Sinclair (1983). The y axis contains the concentration of the variable on an arithmetic or base 10 logarithmic scale and the x axis is plotted on a probability scale, in cumulative frequency percent, equivalent to a range of -2.054 to +2.054 standard deviations about the mean. A straight line plotted on a normal probability scale indicates a dataset having a single, normal population. If the line is curved or plots as several straight lines, this indicates a non-

normal distribution or multiple populations in the data. The cumulative frequency plot can be drawn with point symbols alone, as a line joining the points, or with both point symbols and connecting lines, which can be helpful in identifying gaps in the data showing unique populations. The inflection points on the plot can be read off the concentration axis and used as distinct population breaks for plotting geochemical element distributions on contour maps (Sinclair, 1983).

### PROGRAM DESCRIPTION

The UNISTAT program consists of four modules linked through a main menu system.

#### File Preparation

The datafile has a fixed ASCII format, which can be created with any text editor or with the file preparation module that will allow the user to interactively label variables, and flag missing cases in a data matrix. Datafiles created by this module or text editor can contain up to 40 variables (columns) and 10,000 samples (rows).

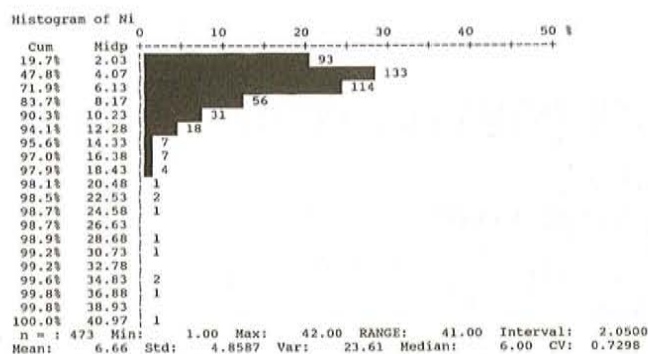
#### Selection

The file and variable selection menu displays all variables contained in the active datafile that must be in UNISTAT internal fixed ASCII format (Table 1). The current variable may be changed or the active file may be changed to display a new set of variables.

**Table 1.** Example dataset showing UNISTAT ASCII file format

Arsenic	ppm	
Copper	ppm	-3 variables(cols)
Nickel	ppm	
179.0	6	11
811.0	3	4
37.2	5	6
34.1	7	7
		-473 samples(rows)





**Figure 1.** Histogram of nickel (ppm) with univariate statistics.

### Univariate Statistics

This is the main core-module of UNISTAT, through which basic descriptive statistics of the variable chosen from the selection menu are displayed and variable subsetting and logarithmic transformation options are processed. All program output is either accomplished or initiated through this module. Graphics output choices include histograms and descriptive statistics having cumulative and frequency percent values displayed on the computer screen or sent to a printer (Figure 1). Cumulative frequency plots can be output to screen or printer with a file option to store the plot, for later plotting on a pen plotter (Figure 2). Printers supported include all Hewlett Packard models, as well as those that are Epson compatible. Only IBM compatible CGA, EGA and VGA graphics adapters are supported for screen displays.

### Plotter Output

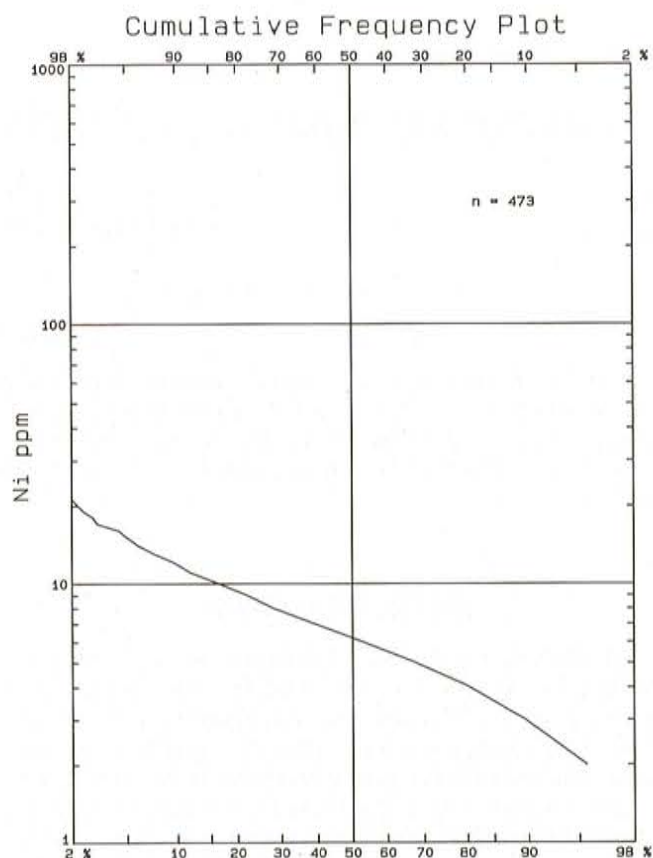
Plot files of cumulative frequency plots created and stored with the univariate statistics module are plotted on a pen plotter with this module. There are various options available including upper-lower data cutoff limits, screen or printer listings of points that fall outside these limits, line types, main title and axis labelling and text- and plot-size scaling, which are all entered through screen menus. Several stored plot files may be plotted on the same graph allowing similar variables or the same variable from different datasets to be plotted together for comparisons. Only plotters using the Hewlett Packard graphics language HPGL are supported.

This program will be distributed as a part of future open file geochemical data releases by the Geochemistry and Geophysics Section. Copies may also be obtained by contacting the author.

## REFERENCES

Hoffman, S.J.

1986: Writing geochemical reports. The Association of Exploration Geochemists, Special Volume No. 12, 29 pages.



**Figure 2.** Cumulative frequency plot of nickel (ppm) plotted on a pen plotter.

Sinclair, A.J.

1981: Applications of probability graphs. In Mineral Exploration. The Association of Exploration Geochemists, Special Volume No.4, 95 pages.

1983: Univariate analysis. In Handbook of Exploration Geochemistry, Volume 2, Statistics and Data Analysis. In Geochemical Prospecting. Edited By R.J. Howarth. Elsevier Scientific Publishing Company, Amsterdam, 1983, pages 59-81.

Stanley, C.R.

1987: Instruction manual for probplot: a computer program to fit mixtures of normal (or log-normal) distributions with maximum likelihood optimization procedures. The Association of Exploration Geochemists, Special Volume No. 14, 39 pages.