GLOSSARY OF STATISTICAL TERMS AND SYMBOLS

Term, Symbol and Formula		Definition
Absolute value $\lvert X \rvert$ = the absolute value of X	$\lvert\ \rvert$	The numerical value of an item regardless of its sign. (Davies, p. 256)
Arithmetic mean $\bar{X} = \frac{\Sigma x}{n}$	$\bar{X}$	The average of a group of items. (Arkin and Colton, p. 11)
Array		Arrangement of numerical data in order of increasing magnitude. (Croxton and Cowden, p. 165)
	β	Probability of accepting a hypothesis when it is false. (Dixon and Massey, p. 80-81)
Bimodal distribution		A frequency distribution with two maxima. (Arkin and Colton, p. 5)
Bimodal distribution $(q + p)^n$		Successive terms of the expansion give probabilities of 0, 1, 2, 3 - - - n, items in a sample of size n, having a characteristic which is found in the proportion p of items in the population from which sample was taken, and is absent in the proportion q, (p+q=1). (Moroney, p. 88-94)
Chi square $\chi^2 = \Sigma \frac{(f_o - f)^2}{f}$	χ^2	Test to determine goodness of fit of observed frequencies to theoretical frequencies; f_o = observed frequency, f = theoretical frequency. (Arkin and Colton, p. 109-l 12)
Class		A subdivision of the observed range of a variable, having stated limits. (Arkin and Colton, p. 2)
Classified data (see frequency distribution)		Numerical data grouped into consecutive classes. (Croxton and Cowden, p. 168)
Class interval	C	Magnitude of the range of values covered by each class. (Arkin and Colton, p. 2)

Term, Symbol and Formula		Definition
Coefficient of variation $V = \frac{\sigma}{\bar{X}}.100$	V	Measure of relative dispersion; σ = standard deviation; X = arithmetic mean (Arkin and Colton, p. 40)
Confidence interval (see Fiducial interval)		The range within which the true value may be expected to fall with a stated probability. (Moroney, p. 238-240)
Contingency table		A table of frequency data arranged under more than one classification. (Davies p. 244)
Correlation		Degree of association between two variables. (Dixon and Massey, p. 3)
Correlation Coefficient $r = \frac{\frac{1}{n}\Sigma(x - \bar{x})(y - \bar{y})}{\sigma x \,.\, \sigma x}$	r	A measure of correlation; $\bar{x}$ and $\bar{y}$ = mean values of x and y, σx and σy = standard deviations of x and y; (limit of $\lvert r \rvert = 1$). (Moroney, p. 286; Croxton and Cowden, p. 931)
Covariance		The expected (mean) value of the product of the deviations of two variables from their respective means. (Davies, p. 246)
Cumulative distribution		An array showing proportion of total greater than, or less than, each recorded value of a variable. (Croxton and Cowden, p. 184)
Degrees of freedom	d.f.	The number of items that are free to vary; if a mean value has been calculated, the value of any item is fixed by the sum of the others, so that d.f. = n − l. (Croxton and Cowden, p. 312)
Dependent variable	Y	The variable whose magnitude is plotted as a function of fixed consecutive values of a second (independent) variable. (Arkin and Colton, p. 4)

Term, Symbol and Formula		Definition
Deviation	d	The difference between an observed value and a standard, which is usually a mean. (Moroney, p 60)
Dispersion		The degree of variation of data around a representative value. (Arkin and Colton, p 8)
e = 2.71828 = limit 1+1/!1+1/!2 +1/!3+---+1/∞	e	The base of the natural, or Naperian, system of logarithms. (Croxton and Cowden, p 924)
$F = \frac{\sigma_1^2}{\sigma_2^2}$	F	A statistic whose distribution measures the significance of the difference between two sample variances, where $\sigma_1 > \sigma_2$ (Dixon and Massey, pp 84-85)
Factorial X! = Xn. Xn_{-1}. Xn_{-2}---1	!	The product of every integral number in a series multiplied together. (Davies, p 257)
Fiducial interval (see Confidence interval)		The interval within which a true value may be said to fall, with a stated probability. (Moroney, pp 238-240)
Frequency	f	The number of items in a specified category, usually a class. (Dixon and Massey, pp 6, 8)
Frequency distribution		A table or graph showing the relative frequencies of items having the various possible values of a specified variable. (Moroney, p 44)
Geometric deviation		Antilogarithm of logarithmic standard deviation. (Shoemaker, and others, p 32)
Geometric mean	Gm	The antilogarithm of the mean of the logarithms of individual values. (Arkin and Colton, p 26)
$Gm = \sqrt[n]{X_1 . X_2 . X_3 - - - - Xn}$		

Term, Symbol and Formula		Definition
Harmonic mean $Hm = \frac{N}{\sum \frac{1}{x}}$	Hm	The reciprocal of the arithmetic mean of reciprocals of individual values. (Croxton and Cowden, p 226)
Histogram		A frequency distribution expressed as a bar chart; width of bar represents class interval, height of bar represents frequency (Moroney, pp 22-23)
Independent variable	X	The variable whose magnitude changes systematically; X also used to denote values of a single random variable. (Arkin and Colton, p 4)
Kurtosis $\pi_4 = \frac{\Sigma X}{n}$	π_4	A measure of the peakedness or flatness of a curve. (Croxton and Cowden, pp 258-259)
Leptokurtic		A narrow, high peaked curve. (Croxton and Cowden, pp 258-259)
Level of significance	α	Probability of rejecting a hypothesis when it is true. (Dixon and Massey, p 80)
Log-normal distribution		A skewed frequency distribution with a mode in the low values, such that the logarithms of the original data yield a normal frequency distribution. (Croxton and Cowden, p 293)
Log-standard deviation $\sigma_L = \sqrt{\frac{\Sigma(\overline{X}_L - X_L)^2}{n-1}}$	σ_L	The standard deviation, expressed as a logarithm, of the logarithms of the original sample values. (Shoemaker and others, p 28)

Term, Symbol and Formula		Definition
Mean deviation $MD = \frac{\Sigma \lvert d \rvert}{n}$	MD	The mean of the absolute values of the deviations of individual items from the group mean. (Arkin and Colton, pp 31-33)
Median		Value of the middle item in an array of numerical data, an average of position. (Arkin and Colton, pp 19-21)
Mode		The most common value, corresponding to the peak of the frequency distribution. (Arkin and Colton, pp 23-27)
	N (n)	Number of items in a sample, sometimes N = number of items in a finite population, where n = number of items in a finite population. (Croxton and Cowden, p 928)
Non-parametric		Independent of the nature of the population distribution. (Dixon and Massey, p 247)
Normal distribution $Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$		A symmetrical bell shaped curve asymptotic to the X axis, - - the normal curve of error. (Dixon and Massey, pp 47-49)
Null hypothesis	Ho	The hypothesis that no significant difference exists between two items that are being compared statistically. (Croxton and Cowden, pp 310-311)
Ogive		A cumulative frequency distribution table, histogram, or curve. (Arkin and Colton, pp 4-5)
Orthogonal		In, experimental design, designed so that for each level of any independent variable, all levels of the other independent variables are represented. (Davies, p 251)

Term, Symbol and Formula		Definition
Parameter		Any measurable characteristic of a sample or a population. (Dixon and Massey, p 33)
Platykurtic		A broad, low peaked curve. (Croxton and Cowden, p 258)
Population	P	1. The entire body of data from which a sample is taken. (Arkin and Colton, p 113) 2. A class or aggregate of objects or events from which a sample is taken. (Krumbein, p 349)
Population mean	μ	The theoretical true mean value, of which the sample mean is an estimate. (Dixon and Massey, p 33)
Probability paper		Graph paper on which cumulative normal frequency distributions plot as straight lines. (Dixon and Massey, p 56)
Probable error $P.E. = 0.6745\sigma$	P.E	The median of the frequency distribution of errors. (Arkin and Colton, p 115)
Quadratic mean $Qm = \sqrt{\frac{\Sigma X^2}{n}}$	Qm	The square root of the mean square of the items in a sample. (Arkin and Colton, p 27)
Random sample		A sample taken in such a way that all items in the sampled population have an equal and independent chance of appearing in it. (Dixon and Massey, p 34)
Range	W	The largest and smallest values in a sample, or the difference between these values. (Arkin and Colton, p 29)
Regression line		The line, or curve from a family of curves, on a scatter diagram, which best fits the empirical relation between a dependent variable and an independent variable. (Arkin and Colton, p 76)

Term, Symbol and Formula	Definition
Residual	That part of the total which is not accounted for by assigned factors; in analysis of variance, the difference between the sum of assigned sums of squares and the total sum of squares. (Dixon and Massey, p 129)
Skewness	The degree of distortion from symmetry exhibited by a curve. (Arkin and Colton, pp 40-41)
Standard deviation σ (s) $\sigma = \sqrt{\frac{\Sigma(d)^2}{n-1}}$	The square root of the variance, σ commonly = population standard deviation, where s = sample standard deviation. (Arkin and Colton, p 33)
Standard error	The standard deviation of a calculated statistical measure. (Arkin and Colton, p 115)
Summation symbol Σ $\Sigma X = X_1 + X_2 + X_3 + \text{---} Xn.$	The sum of variables in a series. (Arkin and Colton, p 206)
Systematic sample	Samples collected at regular pre-determined intervals, such as intersections of a grid. (Krumbein, p 360)
t $t = \frac{\overline{X}_1 - \overline{X}_2}{\sigma\bar{x}_1 - \bar{x}_2}$	The ratio of a statistical measure, normally distributed about a mean of zero, to an estimate of the standard error of that measure. (Croxton and Cowden, p 940)
Trend line	The line or curve on a graph, expressing the best empirical relationship between two variables, commonly a regression line. (Moroney, p 285)
Variance σ^2 (s^2) $\sigma^2 = \frac{\Sigma(d^2)}{n-1}$	The sum of the squared deviations from the mean divided by the degrees of freedom; the standard deviation squared. (Dixon and Massey, pp 19-22)

Term, Symbol and Formula	Definition
Weighted mean	A mean obtained by multiplying each item by a correction factor and dividing the total by the sum of the correction factors. (Arkin and Colton, pp 134-135)
z $z = \frac{1}{2}\log_e F$	The difference between the natural logarithms of two independent estimates of the standard deviation. (Davies, p 262)
Superscript bar —	The mean value ($\overline{X}$), or the estimated value in a population, of a statistical measure calculated from a sample, as σ. (Davies, p 256)

REFERENCES

Arkin, H and Colton, R R, 1955. *Statistical Methods* (Barnes and Noble: New York).

Croxton, F E and Cowden, D J, 1939. *Applied General Statistics* (Prentice-Hall: New York).

Davies, O L, 1949. *Statistical Methods in Research and Production* (Oliver and Boyd: London).

Dixon, W J and Massey, F J Jr, 1951. *Introduction to Statistical Analysis* (McGraw-Hill: New York).

Krumbein, W C, 1960. The geological population as a framework for analysing numerical data in geology, *Liverpool and Manchester Geol. Jour.* (2) 3:341-368.

Moroney, N J, 1953. *Facts from Figures* (Penguin: London).

Shoemaker, E M, 1959. Elemental composition of the sandstone-type deposits, in *Geochemistry and mineralogy of the Colorado Plateau uranium ores*. USGS Prof. Paper 320:25-54.

Till, R, 1974. *Statistical Methods for the Earth Scientist*: An Introduction (Macmillan: London).

1. From Lovering, T G. Glossary of statistical terms used in geological reports. American Geological Institute, Data Sheet 28, by permission.