A few basic statistical procedures are explained for use in describing new or known nematode species. The mean \( \bar{X} \) and the standard deviation \( s \) should be calculated from many specimens (up to and above 30) for every measurement in a sample. Computation of the sample mean (\( \bar{X} \)) and standard deviation (\( s \)) permits an estimate of the confidence intervals to be made for the population mean, \( \bar{X} \pm \frac{(t \times s)}{\sqrt{n}} \), and for a given percentage of the population (\( \bar{X} \pm 2 \times s \)). Observation of several populations is advocated to obtain an estimate of mean and range of a measurement in the species.

**Keywords:** Computation, statistical procedures, descriptions, confidence limits.

Many scientists have expressed concern at the few specimens from which new species are often described: “We must broaden our concept of species by studying more specimens for the variations and the factors that may influence them. …… The description of a new species based on one or two specimens …… differing in one single characteristic in one sex is ridiculous” (Chitwood, 1957). “If the species has sharply distinctive characters 5-10 specimens may be sufficient [if not] at least 20 specimens should be examined and measured.” (Franklin, 1970). “Unfortunately descriptions of new species, …… based on studies of single or very few specimens are still being published.” (Hooper, 1969).

Hooper’s observation still holds, and many new species are described from fewer than ten paratypes. In *Helicotylenchus*, among 83 species described between 1959 and 1971, 30 had ten or less paratypes. The number of specimens studied was not indicated for 12 species. For 65 species described since 1972, 32 had ten or less paratypes. The recording of the mean has shown some progress in recent years, but too many species are still published with only the range (extreme values observed in the sample). As for the standard deviation (S.D.), its use is the exception, rather than the rule in new descriptions. For example, in *Helicotylenchus*, during the period 1959-1971, 62 species were described with only the range, and 21 with range and mean. Since 1972, 30 species were described with range only, and 35 with range and mean. Only one species (*H. depressus* Yeates, 1967) was described including the S.D.

Many taxonomists are probably insufficiently familiar with certain simple statistical concepts to use them with confidence. These are to be found in any
textbook (for example Snedecor & Cochran, 1980), but it may be too difficult to extract from these manuals the few methods appropriate for taxonomic use. The present note is intended to offer a simple, practical account of these methods.

DEFINITIONS

A species should encompass the characteristics of all the individuals which belong to it. These individuals are grouped into populations scattered over the geographical range of the species. The taxonomist considers only one, or in some cases, a few of these populations. As he cannot observe and measure the large number of individuals included in any population, he takes a sample which should be representative of the population, if a sampling procedure has been used that takes a truly random sample from the population. The disparity between the few specimens measured from one locality and the innumerable specimens that constitute the species the sample represents explains why statistical procedures should be followed. The methods explained below require the calculation of two quantities: the mean and the standard deviation. Any textbook on statistics explains the meaning of these parameters and how to calculate them (see paragraphs 3.2 and 3.3 in Snedecor & Cochran, 1980). Most scientific pocket calculators include keys marked "X" for the mean and "S DEV" or "s" for the standard deviation, and may be used for calculating these values, which should be provided for all measurements in all descriptions and redescriptions of species.

STATISTICAL PROCEDURES

Sample mean and population mean: the confidence interval.

The mean \( \bar{X} \), of a measurement as calculated in a sample, gives an estimate of the mean \( \mu \) in the population: \( \mu \) is a fixed value for the population, but \( \bar{X} \) varies from sample to sample. The sample \( \bar{X} \) is the best estimate of the population mean \( \mu \), but it is only an estimate. From the standard deviation \( s \) and the sample size \( n \), a confidence interval \( \bar{X} \pm t \frac{s}{\sqrt{n}} \) (for a normal distribution) can be calculated, where \( n \) is the number of specimens in the sample, and \( t \) is read from a table at the intersection of the line "\( n-1 \)" degrees of freedom (D.F.) and the column "95% probability level" or "5% 2-tail" (see paragraph 4.10 and Table A4 in Snedecor & Cochran, 1980). The value of \( s/\sqrt{n} \) is called the standard error and is a measure of the variability of the population mean.

The interval \( \bar{X} \pm i \) around the mean \( \bar{X} \) observed in the sample has 95% probability of including the true value \( \mu \) of the mean in the population. *)

*) Only in 5% of the cases, will a sample be found whose mean \( \bar{X} \) is too far from the mean \( \mu \) of the population. To reduce this possibility of error, several representative samples from the population can be drawn or measured, or the 99% probability level considered. This makes use of a larger value for \( t \), and will enlarge the confidence interval. The estimate about the value of \( \mu \) will be less precise, but there will only be 1 chance in a 100 of being wrong.