Descriptive Statistics

Prepared by Allan Fraser

October 2016

Application Note: 4

1. Introduction

Knowledge of the descriptive statistics of a data set is a prerequisite for any further statistical analysis of analytical data. This application notes gives some insight into how the mean, degrees of freedom, median and the assessment of kurtosis and skewness is done.

2. Descriptive Statistics

2.1 Degrees of Freedom

Degrees of Freedom = n-1 where n=the number of observations (measurements). Degrees of Freedom abbreviated to '*df*'. Example: we determine the sulphur content of a coal sample by analysis of six replicates of the sample, then, n=6 and df=6-1=5.

2.2. Mean

The mean, or average is calculated as:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n} \qquad [1]$$

Where, n is the number of replicates and x is each individual result.

2.3. Median

The median is the middle number in a sorted list of numbers. Example: find the Median of these concentrations of phosphorous (ppm): [13, 23, 11, 16, 15, 10, 26]. The procedure is to place the numbers in order: [10, 11, 13, 15, 16, 23, 26]. The middle number is 15, so the median is 15. If there are two middle numbers, they are averaged.

2.4. Distribution of Data: Normality, Skewness and Kurtosis

In order to use statistical methods of inference, experimental data do need to be approximately normally distributed. In addition, the kurtosis and skewness of data need to be assessed for significance. Therefore, before any parametric tests are applied all replicate data are examined for normality in their distribution and significance of kurtosis and skewness. The normality in the data distribution is assessed by visual examination of the data in the form of a bar chart. More specifically, the ratio of the mean and median should approximate 1.0 for the data to be normally distributed. In terms of skewness and kurtosis, the standard error of each is compared to their measures. The standard error of kurtosis (S_k) is

$$S_k = \sqrt{\frac{24}{n}} \quad [2]$$

where, n is the number of replicates in a data set.

If the absolute kurtosis is less than three times the standard kurtosis, then the kurtosis is not significant (Quenouille, 1950). The standard error of skewness is:

$$S_k = \sqrt{\frac{6}{n}} \quad [3]$$

If the absolute skewness is less than three times the standard kurtosis, then the skewness is not significant (Quenouille, 1950). The mean, median, kurtosis and skewness are determined with ease using MS Excel's 'data analysis' and 'descriptive statistics' feature.

2.5. Precision

Precision is the closeness with which results of replicate analyses of a sample agree. It is a measure of dispersion or scattering around the mean value and usually expressed in terms of standard deviation. Standard deviation is actually a measure of imprecision, *i.e.* the larger the value the worse the precision. The sample standard deviation for a set of replicate measurements where $n \leq 30$ is given by

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \quad [4]$$

where, *s*, is the sample standard deviation, x_i the individual measurements in the data set, $\bar{x}$ is the mean of the data set at *n-1* degrees of freedom (*df*). For n>30 measurements the formula as in [5] can still be used, as standard deviation, s, approaches the population standard deviation, σ as n increases. In MS Excel, use the formula: =STDEV.S( ) for standard deviation and =STDEV.P( ), for the population standard deviation or σ. The relationship between n, s and σ is shown in the table directly below.

n	σ	s
23	0.101	0.103
53	0.137	0.140
267	0.0784	0.0785
282	0.0764	0.0765

……………………………………………………………………………..

References

Quenouille, M. (1950). *Introductory statistics*. Butterworth-Springer Ltd (1st Edition); 191-192.

Citation

To cite this applications note, use the following:

Fraser A.W. (2016). Descriptive statistics. Application note 4. www.allanfraserandassociates.com and the date that the article was accessed.