SOME APPLICATIONS OF STATISTICS IN ANALYTICAL CHEMISTRY

H.Nascu[a], L.Jäntschi[b], T.Hodisan[b,•], C.Cimpoiu[b], G. Cimpan[b]

[a] *Technical University, Department of Chemistry, 11 C.Daicoviciu, 3400 Cluj-Napoca, Romania*

[b] *"Babes-Bolyai" University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos, 3400 Cluj-Napoca, Romania*

e-mail addresses: ljantschi@hotmail.com; lori@lcibar.utcluj.ro; gcimpan@chem.ubbcluj.ro

CONTENTS

SUMMARY

The review presents some considerations on applications of statistics in analytical chemistry such as the statistics of the point, data analysis by regression, correlation and self correlation, dispersional analysis and ANOVA model, validation of statistic hypothesis.

Statistics have undergone an enormous impact from microelectronics, in the form of microcomputers and hand-held calculators. These have brought difficult statistical procedures within the reach of all practising scientists. The availability of the tremendous computing power naturally makes it all the

• corresponding author

more important that the scientist applies statistical methods rationally and correctly.

1. THE STATISTICS OF THE POINT

1.1 General considerations

Any natural phenomenon and especially those that can be characterized by numerical dates are the result of one or more causes of an action. The experiment, this powerful tool of scientific research, becomes efficient through the skill of the scientist to replace such a complex system of causes with a simple system where only one causative circumstance is allowed to change in time.

The chemist and the physicist have a certain advantage in their research. In their fields of science, experiment has reached a high level of perfection. However, even in these sciences, there are wide possibilities of applying research of a permanent statistical character. We could almost say that in all modern research statistical analysis is used. The elaborate methods aiming to eliminate the effect of circumstances that affect the conditions of measurement, although continually improving, have not reached and could not reach perfection. The scientist himself, and his entire apparatus of observation, constitutes a source of errors too; effects like changes in temperature, humidity, pressure and current, vibration, cannot be completely eliminated.

Statistics has to deal principally with numeric dates, generated by multiple causes. Through making an experiment the scientist wants to solve a certain complex of causes, singling out one by removing all the causes excepting one, or more exactly concentrating his attention on studying one of them by reducing the action of the others as much as possible. Statistics, lacking this possibility, is obliged to analyze data which are influenced by other data and, finally, to determine which are the most important causes and which are the results of the observation that can be attributed to the influence of each cause /1/.

We have to mention that any scientist has to take preventive measures. Thus, the quality of the data has to be examined before jumping to conclusions. This is valid for any kind of data and especially for numeric data with previously verified quality. It is a waste of time trying to apply straight methods of calculation to process primary data that seem to be wrong.

1.2 Localization and scattering indicators

Having a measurement made at more than one time and the results of the measurement divided into two groups of study yields two sets of data. There are two typical fundamental values on which those two sets of data vary statistically:

(1) varying in level, meaning that central value around which the other values gravitate;

(2) varying in the amplitude of spreading observed values around the central value;

The indicators of first type, of level or position of localization are called **means**. Indicators of second type are called **measure scattering indicators**.

There are three forms of mean used more often: *arithmetic mean, median* and *mode*. Means like: *geometric* and *harmonic* are rarely used.

Let the row of measurements be $X_1, X_2, \ldots, X_N$

The **arithmetic mean** (or average) is the number M(X) given by:

$$M(X) = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{1.1}$$

The **median** is the central value of the variable when its values are arranged in the order of their size or that value that has the property that smaller and bigger values appear in equal frequencies. We note X's median as m(X).

Let: $\{1,\ldots,N\}\{1,\ldots,N\}$ be the permutation which arranges the measurements in increasing order:

$$X_{\pi(i)} \leq X_{\pi(i+1)}, \quad i = \overline{1, N-1} \tag{1.2}$$

Then, If N is even,

$$m(X) = (X_{\pi(N/2)} + X_{\pi(1+N/2)})/2 \text{ else } m(X) = X_{\pi((N+1)/2)} \tag{1.3}$$