THEORY OF INFORMATION AS APPLIED TO ANALYTICAL CHEMISTRY. IV.*

EXPRESSION OF THE RESULTS OF QUANTITATIVE CHEMICAL ANALYSES

K.ECKSCHLAGER

Institute of Inorganic Chemistry,
Czechoslovak Academy of Sciences, Prague 6

In the paper attention is called to the lack of uniformity in expressing the results of quantitative analyses. It is suggested to give the most probable result and its uncertainty according to Shannon's relation together with the data on the number of parallel determinations from which the result has been drawn. Tabellated values are presented allowing an easier calculation of the uncertainty from the standard deviation and the range of results.

When expressing the results of quantitative analyses so as to make clear also the accuracy of their determination, we usually give the reliability interval[1] within which the correct result lies with an *a priori* chosen probability, α, denoted as reliability (or significance) level. In practice, we calculate the reliability interval from the estimate of the standard deviation $s = (\sum_{i=1}^{n} (x_i - \bar{x})^2 : (n-1))^{1/2}$ using the relation

$$L_{1,2} = \bar{x} \pm t \cdot s/\sqrt{n}, \qquad (1)$$

where L_1 and L_2 are the lower and the upper limits of the reliability interval, respectively, $\bar{x}$ is the arithmetic mean, t is the Student's coefficient for a chosen α and for the degree of freedom $v = n - 1$ and n is the number of results from which the mean was calculated. For $n \leq 7$ it is more advantageous to carry out the calculation according to Dean and Dixon[2] from the range

$$L_{1,2} = \bar{x} \pm K_n \cdot R, \qquad (2)$$

where R is the range, *i.e.* the difference between the greatest and the smallest result obtained and K_n is a coefficient, known in tabellated form[1,2] for various levels of reliability and for $v = n$.

The choice of the reliability level still remains a subject of discussion: sometimes, particularly when presenting the results of physical measurements[3] α is chosen 0·50 and the value of $(t/\sqrt{n}) \cdot s$ or of $K_n \cdot R$ is then denoted as "probable error" or, with $\alpha = 0\cdot997$ the respective error is called "limit error". To evaluate the precision of results of analyses the chosen values of α are 0·95, 0·99 or 0·997. The choice of the significance level for the calculation of the reliability interval is essential the matter

of convention or arbitrariness. If the chosen value of the reliability level is high (*e.g.* 0·997) the probability that the result lies within an interval will approach certainty, however the interval will be very wide. The results given in this form will be very reliable, but at the same time, of a low precision. In the opposite case, if a lower reliability level is chosen the resulting interval will be narrow, *i.e.* the result will be fairly precise, however the probability that the correct result lies really within this interval is smaller. For $\alpha = 0.50$ this probability is even equal in both cases that the value will lie inside or outside the interval. Moreover, the hitherto employed way of presenting the results of analyses does not make clear about the number of parallel measurements used for calculating the mean value which is an essential factor for the width of the reliability interval. The existing practice of setting down the results of analyses does not all allow to compare their precision as obtained by different authors, owing to the non-uniformity of the choice of α and also because the number of parallel determinations is not given.

Nevertheless, it may be suggested that this number should be given, when presenting the results, even when it was mentioned in the context, namely as a lower index of the bracketed term indicating the precision of the result. This precision, *i.e.* the reliability interval, should not be given for an arbitrary, or some of the commonly used reliability level, but for a level corresponding to the uncertainty of the result. Shannon[4,5] defines the uncertainty of continuously distributed results by the relation

$$P = \int p(x) \cdot \log p(x) \cdot dx , \qquad (3)$$

where $p(x)$ is the so-called probability function. If we substitute for $p(x)$ from the normal mean distribution law

$$p(\bar{x}) = \frac{\exp\left\{-[(\mu - \bar{x})/\sigma_{\bar{x}}]^2/2\right\}}{\sigma_{\bar{x}}\sqrt{2\pi}} , \qquad (4)$$

where the variance of the mean $\sigma_{\bar{x}}^2 = \sigma^2/n$, we obtain for the uncertainty the expression[5,6]

$$P = \sigma_{\bar{x}} \sqrt{2\pi e} . \qquad (5)$$

In the first part of this series[7] we have shown that the uncertainty after analysis may be defined as the width of the reliability interval $(L_2 - L_1)$, as determined for $\alpha = 0.961$. Table I gives the values necessary for the calculation of the reliability interval for $\alpha = 0.961$. If the calculation is carried out using an estimate of the standard deviation s, we employ the column $t/\sqrt{n}$ for $\nu = n - 1$ and if we calculate from the range R, we employ the values from the column K_n for $\nu = n$. No matter which of the mentioned routes of calculation was used, the reliability interval represents the uncertainty of the results in the sense of Shannon's definition. The values $t/\sqrt{n}$ and K_n given in Table I for $\alpha = 0.961$ were calculated using digital computer Tesla 200 as described in papers[7-10].

Table I

Values of $t/\sqrt{n}$ and K_n for $\alpha = 0.961$ and $v = 1$ to 10

v	$t/\sqrt{n}$	K_n	v	$t/\sqrt{n}$	K_n
1	11·510	—	6	0·989	0·429
2	2·833	8·225	7	0·891	0·357
3	1·776	1·487	8	0·818	0·308
4	1·338	0·800	9	0·760	0·272
5	1·127	0·546	10	0·710	0·245

Consequently, the results and their reliability could be reported in such a manner that the uncertainty would be expressed in brackets to distinguish it from the "probable" or "limit" error and with an index denoting the number of parallel determinations from which the mean was calculated as the most probable result.

This way of presenting the results with the indication of their precision and the number of parallel determinations may be illustrated by the two following examples of Ni determination. The determination of Ni in a sample of KCl by means of atomic absorption photometry[11] was carried out three times ($n = 3$), mean $\bar{x} = 2.50 \cdot 10^{-5}\%$ Ni and the estimated standard deviation was $s = 1.3 \cdot 10^{-6}\%$. Distribution of the results is approximately normal as could be stated in a larger checking series of parallel determinations and thus the reliability interval is symmetrical. For $v = n - 1 = 2$ the value of $t/\sqrt{n}$ from Table I is 2·833. The result will be presented as $\bar{x} = (2.50 \pm 0.37)_3 \cdot 10^{-5}\%$ Ni. When determining Ni in KOH by emission spectrography[1] a lognormal distribution was found and therefore the reliability interval is not symmetrical. For $n = 4$ the geometrical mean was found to be $x = 2.9 \cdot 10^{-4}\%$ Ni and the relative value of standard deviations $s_{lg}^+ = 2.42$ for results higher and $s_{lg}^- = 0.408$ for results lower than is the mean. For $v = n - 1 = 3$, when $t/\sqrt{n} = 1.776$ the result will then be presented in the form $\bar{x} = 2.9 \cdot 10^{-4}\%$ Ni $(8.0 \cdot 10^{-5} \div 1.53 \cdot 10^{-3})_4\%$.

REFERENCES

1. Eckschlager K.: *Errors, Measurement and Results in Analytical Chemistry*, p. 97, 104. Published by Státní nakladatelství technické literatury, Prague and Van Nostrand, London 1969.
2. Dean R. B.; Dixon W. J.: Anal. Chem. *23*, 636 (1951).
3. Horák Z.: *Praktická fysika*, p. 52. Published by Státní nakladatelství technické literatury, Prague 1958.
4. Shannon C. E.: *A Mathematical Theory of Communication*. Bell System Technical Journal, 1949.
5. Bell D. A.: *Theorie informace a její použití v technice*, p. 30. Published by Státní nakladatelství technické literatury, Prague 1961.
6. Peters J.: *Einführung in die Allgemeine Informationstheorie*, p. 197. Springer, Berlin 1967.
7. Eckschlager K.: This Journal *36*, 3016 (1971).
8. Lord E.: Biometrika *34*, 41 (1947).
9. David H. A.: Biometrika *38*, 393 (1951).
10. Moore P. G.: Biometrika *44*, 482 (1957).
11. Eckschlager K.: This Journal *34*, 1321 (1969).