How Many Samples Do I need?

Assume you decide that you must be quite certain (say, 95%) that the relative error of the mean of an analysis does not exceed a specific limit. If the upper limit is chosen and the confidence level (C.L.) is chosen, then the number of samples must be adjusted to satisfy both criteria simultaneously. How many samples are needed?

One question immediately arises: Do you know the random error in individual analyses? If not, it will be necessary to do some trial analyses to establish an estimate of that error. Then, look at the equation

$$\mu = \bar{X} + t(\sigma / \sqrt{n})$$

when $\sigma$ is unknown

Our task is to find an $n$ so that with the method’s $s$, the confidence limit is some fraction of the mean ($\bar{X}$). Now, let the maximum allowable relative error be represented by $R$. Then

$$t \frac{s}{\sqrt{n}} = R\bar{X}$$

From this formula, we find.

$$n = \frac{t^2 s^2}{R^2 \bar{X}^2}$$

but $s / \bar{X}$ is the actual (found) relative standard deviation while $R$ in the equation is the desired relative standard deviation. We can rewrite equation-1 as

$$n = t^2 \frac{s^2}{R^2 \bar{X}^2} = t^2 \left(\frac{R_{\text{actual}}}{R_{\text{desired}}}\right)^2$$

Solving for $n$ is not straightforward, however, because the value of $t$ itself depends on $n$, the number of samples. As a result, the equation is solved by iteration. That means the answer from a trial is placed back into the equation repeatedly until the answer does not change. Here, the most efficient way to work is to use the $t$ for the 95% C.L. but for an infinite number of samples that is, $t = 1.96$. $n$ is found, rounded to the next highest integer, and substituted into the equation. A new $n$ is found, and so forth.

EXAMPLE-1

Suppose a method for the determination of boric acid in solution for production of eyedrops gives $3.0 \pm 0.17$ mg mL$^{-1}$ boric acid. That is, $s = 0.17$ mg mL$^{-1}$. If the process requires that the solution be within 5% of the stated concentrations, how many trials should be done on each batch to ensure that the prescribed limits are attained with a 95% confidence level?

SOLUTION
From the data, $R = 0.05$, $s = 0.17$, $X = 3.00$.

Begin with $t$ for an infinite number of samples, which for the 95% C.L. is 1.96

$$n = t^2 \frac{s^2}{R^2 X^2} = (1.96)^2 \frac{(0.17)^2}{(0.05)^2(3.00)^2} = 3.8 \times 1.28 = 4.9$$

rounded to 5 samples

For $n = 5$, $t = 2.78$ at the 95% C.L. This value of $t$ is substituted and

$$n = (2.78)^2 \times 1.28 = 7.7 \times 1.28 = 9.9$$

rounded to 10 samples

For $n = 10$, $t = 2.26$, which gives $n = 7$. The next iteration has $n = 8$, followed another iteration with $n = 7$.

To obtain the result with the desired confidence, eight samples should be run. Clearly, if this is to be a routine analysis, some time should be spent improving the precision of the analytical methodology.

EXAMPLE-2

1. An analyst is asked to determine lead in a consignment of fruit juice. The client specifies that the lead content is of the order of 100 mg/kg (ppb) and that he requires an “accuracy” of 5 mg/kg and accepts a 95% confidence level. Calculate the sample size necessary to satisfy this request assuming that, at the specified concentration level, the precision of the analytical method used is known to be $\pm 8$ mg/kg (注：这里假设方法的总体标准为已知).

SOLUTION

The precision of the analytical method used is known to be $\pm 8 \mu g/kg$, that means $\sigma = 8 \mu g/kg$

From a $t$ table it can be found that $t_{0.05, \infty} = 1.96$

The client’s request for a 95% confidence level means that $\mu$ lies in the range

$$\bar{X} \pm \frac{1.96\sigma}{\sqrt{n}}$$

where $\frac{1.96\sigma}{\sqrt{n}} = 5$

It follows that: $\sqrt{n} = \frac{1.968}{5} = 3.14$, $n \sim 10$

Ten samples must be analyzed to meet the client specifications; note, however, that the underlying assumption in the above approach is that the analytical method used is free from systematic error.