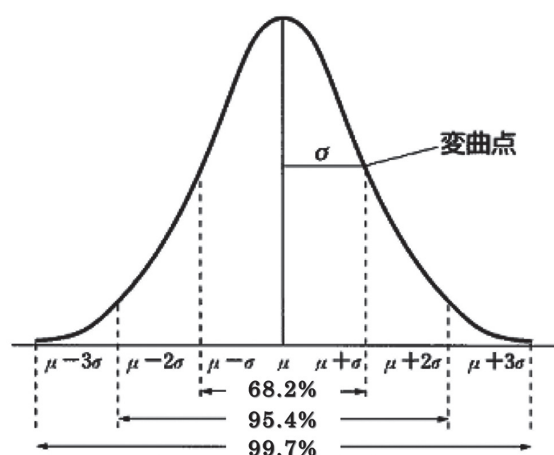


分析能パラメータの併行精度は標準偏差や相対標準偏差で評価するものであり、基礎的な統計解析手法が分析法バリデーション実施のため必要となります。本章では、必要な統計解析について簡単に触れることにします。

1. 正規分布と標準偏差、相対標準偏差

われわれの測定データは、様々な原因によってばらつきます。ばらつくということは、データが分布をもっているということです。分析を無限に行ったときの測定データの分布は、図1のような曲線(ガウスの誤差曲線:正規分布)になり、バラツキを標準偏差 σ (シグマ)を用いて平均値の両側に $\pm\sigma$ だけの幅をとると、この区間に入る確率は68.2%であり、 $\pm2\sigma$ をとれば95.4%、 $\pm3\sigma$ をとれば99.7%(1000回測定した場合、この $\pm3\sigma$ の範囲からはずれの可能性は3回あります:千三つの法則)であることが統計的計算により求められています。



(図1) 正規分布

われわれは、分析法バリデーションで測定する5~9個のデータからガウス分布であらわされる母集団の分布の姿、すなわち分布の中心位置(母平均 μ)と分布のバラツキ(母標準偏差 σ)を推定します。測定データから得られる計算値(実測値の平均 $\bar{X}$ 、不偏分散 V)を統計量といい、次のように表します。たとえば、23.600、23.783、23.583、23.617、23.610、23.639というデータがあれば、

$$\bar{X} = 1/6 (23.600 + 23.783 + 23.583 + 23.617 + 23.610 + 23.630) = 23.637$$

個々のデータの平均値からの差(偏差)の2乗の和を偏差平方和といいます。これは、2乗することにより、負の差もバラツキとして正の数値として取り扱うことができるからです。

すなわち、2乗しなければ、

$$(23.600 - 23.637) + (23.783 - 23.637) + (23.583 - 23.637) + (23.617 - 23.637) + (23.610 - 23.637) + (23.630 - 23.637) = (-0.037) + (+0.146) + (-0.054) + (-0.020) + (-0.027) + (-0.007) \approx 0$$

となり正の数値と負の数値が打ち消し合ってほぼゼロとなります(図2)が、

表 10 視覚的評価に基づく方法による定量限界の評価

0.005 % 相当	検出できる	検出限界
0.006 % 相当	検出できる	
0.01 % 相当	定量できる	定量限界
≥ 0.01 % 相当	直線性、真度、精度が良	

②シグナル対ノイズに基づく方法

この手法は、検出限界と同様、ベースラインノイズを伴う分析法にのみ適用できます。シグナル対ノイズ比は、分析対象物を既知の低濃度で含有する試料のシグナルをブランク試料のシグナルと比較することによって求めることができます。これを用いて分析対象物が確実に定量できる最低の濃度を求めます。定量限界設定のための標準的なシグナル対ノイズ比は、10 : 1 です。

③レスポンスの標準偏差と検量線の傾きに基づく方法

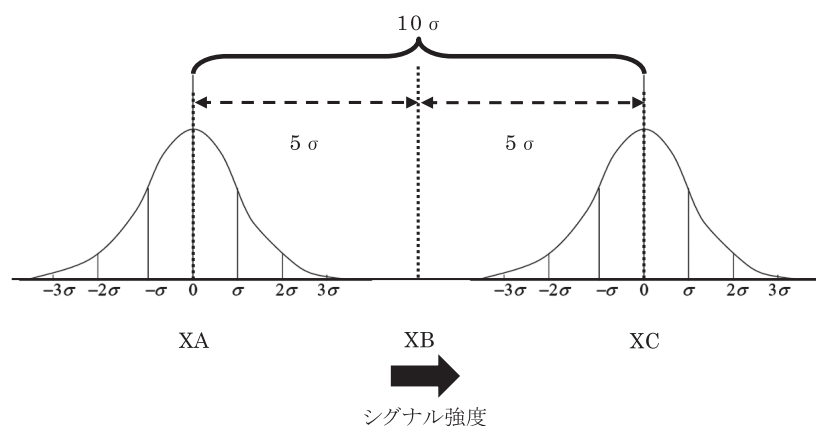
定量限界(QL)を次式により決定することもできます。

$$QL = 10 \sigma / S$$

ここで、 σ : レスポンスの標準偏差

S : 検量線の傾き

傾き S は、分析対象物の検量線から推定できます。標準偏差 σ については、種々の推定方法があるが、検出限界と同様、検量線に基づく方法が一般的です。検量線に基づく方法とは、定量限界付近の濃度の分析対象物を含む試料を用いて、検量線を検討するものです。回帰直線の残差の標準偏差又は回帰直線から推定した濃度ゼロにおけるシグナルの標準偏差を標準偏差 σ として利用できます。検出限界の場合 3.3σ の強度(図 21)としましたが、定量限界は X_B のシグナルの強度を X_A から 5σ 離し、 X_C のシグナルの強度を X_B から 5σ 離すようにします。従って、定量限界のシグナル強度は $5\sigma + 5\sigma = 10\sigma$ となります(図 22)。

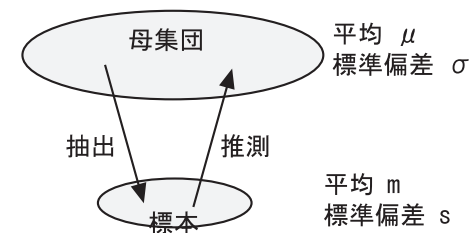


(図 22) 定量限界の求め方(レスポンスの標準偏差と検量線の傾きに基づく方法)

●統計解析に関する Q&A

Q 母集団と統計量とは？

データ分析において、その対象とする全ての要素が含まれる集合のことを「母集団」といいます。母集団全体を分析できればよいのですが、規模が大き過ぎる等して、実施が難しいような場合には、母集団全体ではなく、その一部分からデータ分析をしなければならないこともあります。この母集団から抽出された一部分のことを「標本」といいます。我々が母集団から得たい情報は、母集団の分布の姿、すなわち分布の中心位置(母平均) μ 、分布のパラツキ(母分散)、母標準偏差 σ です。そのために、標本を分析し、得られた分析値から推測します。分析値から計算した値を統計量といいます。統計量としては、実測値の平均値(m)、偏差平方和、不偏分散、標準偏差(s)等があります。



(図 41)

Q 母平均および母分散の区間推定の意義は？

統計量により我々が求めた推定値(母平均および母標準偏差)には、その値がどの程度信頼できるかの情報が含まれていません。そこで、推定値を元に、母集団の値が、どの程度の信頼度で、どの範囲に含まれるかを推定するのが、区間推定です。詳しくは、第1章の4を参照ください。

Q 区間推定のために用いる統計表は？

母平均の区間推定には、母標準偏差が分かっている場合は正規分布表、分かっていないときはt分布表(概ねn数が30未満のとき)を用います。母標準偏差の区間推定には χ^2 分布表を用います。母平均の信頼区間は次式で表されます。

$$X \pm t(n-1, \alpha) \times \sqrt{(V/n)}$$

Xは平均値、 $t(n-1, \alpha)$ はt表から求めた数値、Vは不偏分散、nは標本数
母標準偏差の信頼区間は次に示すとおりです。

$$\frac{(標本数-1) \times 不偏分散}{\chi^2_{上側}} \sim \frac{(標本数-1) \times 不偏分散}{\chi^2_{下側}}$$

(図 42)

χ^2 は χ^2 分布表から読み取った数値です。