データ評価のための統計的方法

——分散分析の利用——

田　中　秀　幸

1　はじめに

第1回，第2回と統計の基礎から正規分布を用いた推定・検定の解説を行ってきた。これらの手法は統計的手法の中でも根本となる部分であり，統計的手法の応用もすべてこの基礎から始まる。

今回解説するのは統計的応用法のなかの分散分析である。分散分析は分析化学におけるデータ解析に特によく用いられる手法であるので，よく理解してほしい。

2　分散分析とは

データを取得した際に，そのデータに複数のばらつきの要因が含まれていることがある。この複数のばらつきを分離する手法が分散分析である。

データに複数のばらつきの要因が含まれるとは，例えばあるロットの溶液から瓶に小分けしたときの瓶間の濃度のばらつきと繰り返しの濃度のばらつきが測定データに含まれる場合や，ある測定結果の日間変動と日内変動がデータに含まれる場合などである。このようなデータから，それぞれのばらつきの大きさを推定する方法が分散分析である。

まず，分散分析の概念を紹介するための例として，表1に標準物質の測定データを示す。これは，二つの標準物質を5回ずつ繰り返し測定した結果である。

表1では平均値が0.3異なっているので，瓶Aと瓶Bの濃度が異なっている，と考えられそうである。

では，表2ではどうだろうか。この場合も表1と両者の平均値は等しいので二つの瓶の間で差があるように見える。しかし，個別のデータを見てみると，表1では各瓶の繰り返しデータの最大値と最小値の差が0.3ほどであるが，表2では6近くとなる。表2では，平均値の差が表1と等しいが，しかし平均値の差である0.3という値は瓶Aの濃度と瓶Bの濃度が異なることが原因となっているわけではなく，繰り返しのばらつきによって偶然的に引き起こされたものと考えられる。よって表2では，瓶Aと瓶Bの間に濃度の差があるとはいえないだろう。つまり，瓶Aと瓶Bの差があるかどうかを知りたければ平均値の差だけを比べていたのではわからない。この平均値の差と，繰り返しのばらつきの大きさを比べて総合的に判断しなければならないのである。このとき指標として用いられるのが分散である。つまり，瓶間の濃度の違いの分散と繰り返しの分散を比較するという手法を用いて評価する。

表1　瓶Aと瓶Bの違い　その1（単位：mg/L）

	1	2	3	4	5(回)	平均
瓶A	99.9	100.2	100.1	100.2	100.0	100.08
瓶B	100.2	100.5	100.3	100.4	100.5	100.38

表2　瓶Aと瓶Bの違い　その2（単位：mg/L）

	1	2	3	4	5(回)	平均
瓶A	101.9	99.0	103.6	98.2	97.7	100.08
瓶B	100.3	98.5	102.1	97.9	103.1	100.38

3　分散分析の構造

本章では分散分析法の原理とそのデータ構造について解説する。ここでは一番単純な分散分析を考える。一番単純な分散分析とは，ばらつきの要因が測定の繰り返しのほかに要因が一つだけ含まれる場合である。このようなデータに対する分散分析を一元配置の分散分析という。もし，繰り返し以外の要因がさらに増えると二元配置，三元配置もしくは多元配置の分散分析という。本解説では多元配置の分散分析については解説しないが，多元配置の分散分析も基本的には一元配置の分散分析の拡張である。一元配置と多元配置の分散分析の違いはあとで紹介する。

一元配置の分散分析を適用する方法を解説するために以下の例を用いる。

例：標準物質を1回に大量に作製し，それを小分けして瓶詰めを行った。瓶詰めされた標準物質間に濃度の差があるのかどうか調べたい。このとき，瓶詰めされた標準物質から5個瓶を取り出し，それぞれの瓶の標準

Fundamental Knowledge for Reliable Analysis—Statistical Methods for Data Analysis: Part 3.

表 3　濃度測定結果（単位：mg/L）

瓶名＼繰り返し	1	2	3(回)
瓶 A	100.2	100.3	100.0
瓶 B	99.8	99.9	99.7
瓶 C	100.3	100.4	100.2
瓶 D	100.0	100.1	100.0
瓶 E	99.7	99.8	99.9

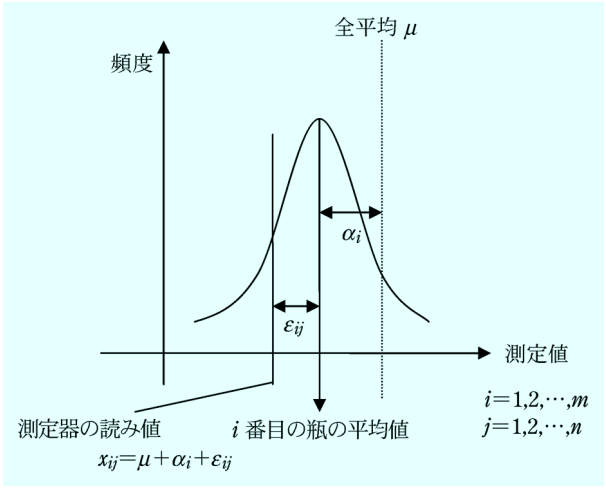

図 1　データの構造

物質を 3 回の繰り返し測定を行って，その結果から瓶の間に濃度の差があるかどうか，またその濃度の差はどの程度であるのかを調べる。

実験を行った結果，表 3 を得た。このような場合に分散分析法は非常に有用である。図 1 に表 3 で示されたデータの構造を示す。

データが構造を持つということを式で表すと，この標準物質は 1 回に大量に作製されているので，この大量に作製された標準物質の濃度の真値が存在すると考えられる。そして，その標準物質を瓶詰めすれば，ある瓶ではその大量にあった標準物質の濃い部分が詰められたかもしれない，もしくは，薄い部分が詰められたかもしれない。つまり，瓶によって，何らかの値のかたよりが存在するはずである。また，繰り返し測定を行っているが，その値は測定ごとにばらついている。つまり，一つ一つのデータにも何らかの値の変動が含まれている。これを式に表すと，

$$x_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad \cdots\cdots(1)$$

となる。つまり，小分けされた標準物質が入っている i 番目の瓶（$i = 1, \cdots, m$）を一つ取り出し，それを繰り返し測定した j 番目の結果（$j = 1, \cdots, n$）を x_{ij}，その標準物質の真の濃度を μ，i 番目の瓶に入っている標準物質の濃度と μ との差を α_i，繰り返しの変動を ε_{ij} としている，ということである。

モデル式が式(1)で表されるということは，分散分析を行うことができる前提の最重要部である。このモデルで表すことができるということとともに，以下の前提を満たすことが分散分析法を適用する条件となる。

1）誤差の不偏性
2）誤差の等分散性
3）誤差の独立性
4）実験のランダム化

誤差の不偏性とは ε_{ij} がもつ期待値が 0 である，ということを表す。つまり，測定ごとに繰り返しのばらつきは存在するが，その繰り返しのばらつきを無限個集め平均値をとると 0 になるということである。

誤差の等分散性とは，この例でいうと各瓶における繰り返し測定のばらつきを表す母分散がすべて等しいと考えられる，ということである。ある瓶では繰り返しのばらつきが非常に大きく，ある瓶では非常に小さい，というときには分散分析は使えない。

誤差の独立性とは，繰り返しのばらつきと瓶間のばらつきは独立である，ということを表している。つまり，瓶によって繰り返しのばらつきに何らかの傾向があってはいけないということである。

実験のランダム化とは，連載第 1 回で解説した実験の順番をランダムに行う，ということである。これは 1）2）3）の条件と非常に関係が深い。つまり，第 1 回でも解説したようにランダム化を行わないと他のばらつきの要因が不意に入り込み，分離できなくなる事態が起こる。分散分析を行う際は特に実験計画を入念に練る必要がある。

ここで，今回のデータは先に挙げた前提をすべて満たしているとしたとき，x の変動（二乗和）を考えると，式(2)となる。

$$S_T = \sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2 \quad \cdots\cdots(2)$$

これは各データが全平均（i, j 関係なくすべてのデータの平均値）からどの程度離れているかというものの指標である。これを全変動と呼ぶ。先ほどの前提条件を満たしていると，式(2)を二つの変動に分解することができる。

$$\begin{aligned} S_T &= \sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2 \\ &= \sum_i \sum_j (\bar{x}_i - \bar{\bar{x}})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 \\ &\qquad \cdots\cdots(3)^{*1} \end{aligned}$$

式(3)の左辺第 1 項は各瓶の平均値と全平均との差から求められる変動で，第 2 項は各瓶の測定値とその瓶の平均値との差から求められる変動，つまり繰り返しの変動を表している。これを，

*1　左辺を $\sum_i \sum_j \{(x_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{\bar{x}})\}^2$ と変形し，展開すると $\sum_i \sum_j (x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{\bar{x}})$ という項がでてくるが，前提を満たしていればこの項を計算すると 0 となる。

$$S_A = \sum_i \sum_j (\bar{x}_i - \bar{\bar{x}})^2 \cdots\cdots(4)$$
$$S_e = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 \cdots\cdots(5)$$

とする。そうすると式(6)が成り立つ。

$$S_T = S_A + S_e \cdots\cdots(6)$$

つまり，分散分析を行うことができる前提が成立するならば，全変動（S_T）を級間変動 S_A（ここでは瓶間の変動）と級内変動 S_e（ここでは繰り返しによる変動）とに分解できるということである。

次に自由度を考える。自由度は本連載第1回で解説したように，データ数から用いられる平均値の個数を引いたもので求められる。T の自由度は，データ数が mn 個，平均値が全平均一つであるので，$mn-1$ となる。A の自由度は，データ数 m 個（各瓶の平均値の個数），平均値は全平均一つであるので，$m-1$ となる。e の自由度は，データ数 mn 個，平均値の個数は，各瓶の平均値 m 個であるので，$mn-m=m(n-1)$ となる。

これらの結果より，自由度も変動と同じく分解できていることがわかる。つまり，

$$mn-1=(m-1)+m(n-1),\ f_T=f_A+f_e \cdots\cdots(7)$$

が成立する。ここで，f は自由度を表す。

これで，各変動と各自由度が求まった。よって，変動を自由度で割れば分散が算出できる。この結果を一覧にしたものを表4に示す。

通常，分散分析した結果は表4のような分散分析表として表される。統計解析ソフト等を用いて分散分析を行った際も，一番右の欄の分散の期待値というところを除いては同様の表が計算されるはずである。分散の期待値についてはこのあと解説する。

次に，ここで算出された V_A と V_e はいったい何を推定している分散なのかということについて考えよう。普通に考えると，V_A は瓶間の分散を n 倍[*2] したものに対応していそうな気がするが，詳細に見てみよう。

級間変動の算出式，式(4)に，分散分析のモデル，式(1)を代入すると，

$$\begin{aligned} S_A &= \sum_i \sum_j (\bar{x}_i - \bar{\bar{x}})^2 \\ &= \sum_i \sum_j \{(\mu + \alpha_i + \bar{\varepsilon}_i) - (\mu + \bar{a} + \bar{\bar{\varepsilon}})\}^2 \\ &= \sum_i \sum_j \{(\alpha_i - \bar{\alpha}) + (\bar{\varepsilon}_i - \bar{\bar{\varepsilon}})\}^2 \end{aligned}$$

となり，分散分析の前提である誤差の独立性が成立しているとすると，上式は，

$$S_A = \sum_i \sum_j (\alpha_i - \bar{\alpha})^2 + \sum_i \sum_j (\bar{\varepsilon}_i - \bar{\bar{\varepsilon}})^2 \cdots\cdots(8)$$

となる。同様に S_e について考えると，

$$\begin{aligned} S_e &= \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 \\ &= \sum_i \sum_j \{(\mu + \alpha_i + \varepsilon_{ij}) - (\mu + \alpha_i + \bar{\varepsilon}_i)\}^2 \\ S_e &= \sum_i \sum_j (\varepsilon_{ij} - \bar{\varepsilon}_i)^2 \cdots\cdots(9) \end{aligned}$$

となる。次に，この S_A, S_e が何を推定しているのかを求めるため，連載第1回目で解説した期待値を計算する。

$$E(S_A) = E\left[\sum_i \sum_j (\alpha_i - \bar{\alpha})^2 + \sum_i \sum_j (\bar{\varepsilon}_i - \bar{\bar{\varepsilon}})^2\right]$$

$$E(S_A) = nE\left[\sum_i (\alpha_i - \bar{\alpha})^2 + \sum_i (\bar{\varepsilon}_i - \bar{\bar{\varepsilon}})^2\right] \cdots\cdots(10)$$

ここで，

$$\sigma_A^2 = E\left[\frac{\sum_i (\alpha_i - \bar{\alpha})^2}{m-1}\right],$$
$$\sigma_e^2 = E\left[\frac{\sum_i (\varepsilon_{ij} - \bar{\bar{\varepsilon}})^2}{mn-1}\right],\ V(\bar{\varepsilon}_i) = \frac{\sigma_e^2}{n} \cdots\cdots(11)$$

であるとする。式(11)の意味するところは，A の影響による母分散を σ_A^2 とし，繰り返しの影響による母分散を σ_e^2 とした，ということである。また，連載第1回目で解説したように，平均値の分散はデータの分散をデータの個数で割ったものと等しくなる。

式(11)を式(10)に代入すると，

$$E(S_A) = n(m-1)\sigma_A^2 + n(m-1)\frac{\sigma_e^2}{n}$$
$$E(S_A) = n(m-1)\sigma_A^2 + (m-1)\sigma_e^2 \cdots\cdots(12)$$

となる。また同様に式(9)の期待値を取ると，

$$E(S_e) = E\left\{\sum_i \sum_j (\varepsilon_{ij} - \bar{\varepsilon}_i)^2\right\} \cdots\cdots(13)$$

となる。ここで，

$$E\left\{\frac{\sum_j (\varepsilon_{ij} - \bar{\varepsilon}_i)^2}{n-1}\right\} = \sigma_e^2 \cdots\cdots(14)$$

とする。これは，ある瓶内で繰り返しを n 回行い，分散を算出したとすると，この分散は誤差の等分散性より繰り返しの母分散の推定値であるということを意味す

表4　一元配置の分散分析表

要因	S(変動)	f (自由度)	V(分散)	$E(V)$ (分散の期待値)
A	$S_A=\sum_i\sum_j(\bar{x}_i-\bar{\bar{x}})^2$	$m-1$	$V_A=S_A/f_A$	$E(V_A)=\sigma_e^2+n\sigma_A^2$
e	$S_e=\sum_i\sum_j(x_{ij}-\bar{x}_i)^2$	$m(n-1)$	$V_e=S_e/f_e$	$E(V_e)=\sigma_e^2$
T	$S_T=\sum_i\sum_j(x_{ij}-\bar{\bar{x}})^2$	$mn-1$		

[*2] なぜ n 倍なのかというと，通常の分散の算出では，$\sum_i (\bar{x}_i - \bar{\bar{x}})^2$ でよいが，さらに j の分も和をとっている，つまり n 倍しているからである。全変動の分解という観点から j の分の和も必要となる。

表 5　分散分析結果

要　因	二乗和 S	自由度 f	分散 V	分散の期待値 $E(V)$
瓶	0.5907	4	0.1477	$\sigma_e^2+3\sigma_A^2$
繰り返し	0.1133	10	0.01133	σ_e^2
合　計	0.7040	14		

る。式(14)を式(13)に代入すると，

$$E(S_e) = E\left\{\sum_i\sum_j(\varepsilon_{ij}-\bar{\varepsilon}_i)^2\right\} = \sum_i(n-1)\sigma_e^2$$
$$E(S_e) = m(n-1)\sigma_e^2 \cdots\cdots(15)$$

となる。式(12)，(15)の結果を各自由度で割れば，各分散が何を推定しているのかということがわかる。

$$E(V_A) = \sigma_e^2 + n\sigma_A^2 \cdots\cdots(16)$$
$$E(V_e) = \sigma_e^2 \cdots\cdots(17)$$

つまり，分散分析で算出された瓶間の分散は瓶間の母分散をそのまま推定しているわけではなく繰り返しのばらつきを含んでいる。また，繰り返しの分散は繰り返しの母分散をそのまま推定している。表 4 の分散の期待値の欄は上記のように計算されたものである。

よって，瓶間の母分散の推定値として用いることができるのは，

$$\hat{\sigma}_A^2 = \frac{V_A - V_e}{n} \cdots\cdots(18)$$

である。また，繰り返しの母分散の推定値は，

$$\hat{\sigma}_e^2 = V_e \cdots\cdots(19)$$

である。このようにして複数のばらつきの要因を分解しそれぞれの分散の推定値を求める。

では先ほどの例に戻って，表 3 の結果に分散分析を適用してみる。表 5 は表 3 を分散分析した結果である。

要因「瓶」の分散は 0.1477 であるが，それは，$\sigma_e^2 + 3\sigma_A^2$ の推定値として求められたものである。つまり，「瓶」の分散の欄に書かれている数字は，繰り返しの分散一つと瓶間の濃度の違いの分散三つ分が足されたものが算出されている。同様に，「繰り返し」の分散の欄の 0.01133 は繰り返しのばらつきの分散一つ分の推定値が算出されている。つまり，「繰り返し」の分散はそのまま繰り返しの分散の推定値として考えられるということである。

よって，

$$\hat{\sigma}_A = \sqrt{\frac{0.1477 - 0.01133}{3}} = 0.2132 \cdots\cdots(20)$$

$$\hat{\sigma}_e = \sqrt{0.01133} = 0.1065 \cdots\cdots(21)$$

となり，瓶間のばらつきの標準偏差の推定値と，繰り返しのばらつきの標準偏差の推定値を求めることができた。

4　分散分析を用いた検定

第 3 章で示した方法によってばらつきを分離することができる。しかし実際には，ばらつきを分離することが目的でなく，瓶間のばらつきが存在するのか，しないのか，ということを判定したいということが多々ある。このときには分散分析を用いた検定を行う。ただし，検定を行うためには先ほどあげた分散分析を適用するための前提条件以外にもう一つ前提条件が必要となる。それは，

5）誤差の正規性

である。誤差の正規性とは繰り返しのばらつきの確率分布が正規分布に従っているということである。検定を行う際には正規分布の性質を用いて行うので，この前提が必要となる。

分散分析を行った結果を用い，瓶間の濃度の違いが本当にあるのかどうか，繰り返しのばらつきと大きさを比べることによって調べることができる。今回の例では，$\sigma_e^2 + 3\sigma_A^2$ と σ_e^2 の推定値の大きさを比べる。つまり，V_A と V_e の大きさを比べると，V_A には繰り返しの分散のほかに瓶間の濃度の違いの分散が入っている。もし，本当に瓶によって濃度の違いがないのであれば，$\sigma_A^2 = 0$ と考えられるので，どちらの分散も繰り返しの分散一つ分である σ_e^2 が推定されているはずである。つまり両者の値はほぼ等しくなければならない。また，瓶によって濃度の違いが大きくあるのであれば，$\sigma_A^2 > 0$ となり，V_A は V_e より大きくなるだろう。よって，V_A と V_e の大きさを比べることによって，瓶間に本当に濃度の違いがあるのかどうかが判定できる。

実際の判定方法だが，次のような値を考える。

$$F = \frac{V_A}{V_e} \Longrightarrow \frac{\sigma_e^2 + 3\sigma_A^2}{\sigma_e^2} \cdots\cdots(22)$$

瓶間の濃度差がないのであれば，F の値は 1 に近づき，瓶間の濃度差が大きいのであれば，F の値は 1 よりはるかに大きくなる。ではどのくらい大きければ瓶間の濃度差が存在すると判定されるのだろうか。これについては，F の値がどのようになれば差が存在するか，ということが十分に調べられ，その値が数値表として統計の教科書に載っている。その表の一部を表 6 に示す。

この表の使い方であるが，一番左の列に書かれている 1～10 の数字は V_e の自由度である。今回の例では，分散分析表（表 5）を見ると 10 である。一番上の列の 1～10 の数字は，V_A の自由度である。今回の例では 4 である。よって，縦の 10 番目，横の 4 番目の欄を見ると，3.48 とある。つまり，F の値が 3.48 より大きければ瓶間の濃度差が存在する，ということになる。この表に書かれている数字のことを F 境界値と呼ぶ。

表 6　F 分布表（5%）

$V_e \diagdown V_A$	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	237	239	241	242	244
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98

詳しく表を見てみると，自由度が少なければ F 境界値も大きい。自由度は測定の回数で決まってくる値である。つまり，ほんの数回しか測っていない場合は，たまたま繰り返しのばらつきが小さく出てしまう場合がある。しかし，F 境界値の値が大きいので，そのような場合でも簡単には瓶間の濃度差がある，とは判定されない。また，測定回数が多くなれば，F 境界値は小さくなる。つまり，求められた分散の精度が上がっているので，「瓶」の分散がある程度「繰り返し」の分散より大きくなれば瓶間の濃度差が存在すると判定されるのである。

では，計算してみよう。この実験での F の値は，

$$F = \frac{0.1477}{0.01133} = 13.03 \quad \cdots\cdots(23)$$

である。また，今回の場合 F 分布表から F 境界値は 3.48 であることが分かる。よって，

$$13.03 > 3.48 \quad \cdots\cdots(24)$$

であるので，瓶間の濃度差が存在する，ということが分かった。ここで紹介した検定法を F 検定という。

5　標準物質への値付けのばらつきの大きさ

通常標準物質では，今回の例のような手法によって全平均を求め，その全平均をここで生産された瓶詰めの標準物質の値として採用する。では，この瓶詰めされた標準物質の濃度のばらつきはどのような大きさになるのか考える。

この標準物質の濃度は，全平均である，

$$\bar{\bar{x}} = 100.02 \text{ mg/L} \quad \cdots\cdots(25)$$

によって推定される。この値がどのような構造を持っているのか考えると以下のようになる。

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}}{mn} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} (\mu + \alpha_i + \varepsilon_{ij})}{mn}$$
$$= \mu + \frac{\sum_{i=1}^{m} \alpha_i}{m} + \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} \varepsilon_{ij}}{mn} \quad \cdots\cdots(26)$$

ここで，m は測定を行った瓶の数（今回は 5）であり，n は各瓶での繰り返し回数（今回は 3）である。つまり，完全に標準物質の濃度を推定できているわけではなく，一部瓶間の濃度差と，繰り返しのばらつきが含まれている。よって，その部分を測定結果の曖昧さとして見積もらなければならない。

では，瓶間の濃度差はどの程度ばらつきとして含まれるのかを見てみると，m 個の瓶の平均値分だけずれている。よって，m 個の平均値の分散がばらつきに含まれるということである。つまり，標準物質の濃度を決定する際のばらつきの要因として，瓶間の濃度差が存在し，そのばらつきの大きさを表す標準偏差は，

$$S_A = \sqrt{\frac{\hat{\sigma}_A^{\,2}}{m}} = \sqrt{0.04544/5} = 0.09534 \text{ mg/L} \quad \cdots\cdots(27)$$

である。一方，繰り返しの不確かさだが，繰り返しのばらつきは mn 個のデータの平均値になっていることから，

$$S_e = \sqrt{\frac{\hat{\sigma}_e^{\,2}}{mn}} = \sqrt{\frac{0.01133}{5 \times 3}} = 0.02749 \text{ mg/L} \quad \cdots\cdots(28)$$

となる。その他にばらつきの要因がないとするならば，この標準物質の濃度のばらつきは，

$$S_c = \sqrt{0.09534^2 + 0.02749^2} = 0.09922 \approx 0.10 \text{ mg/L} \quad \cdots\cdots(29)$$

となる。

では，この標準物質に値とばらつきをつけたので，売り出そうと思う。そのとき，各標準溶液が入った瓶に認証書を添付するが，そこにはどのように書けばよいのだろうか。

溶液の濃度：100.02 mg/L，ただし溶液の濃度は標準偏差 0.10 mg/L で表されるばらつきを持つ。

これでよいのだろうか？　これでは非常に大きな問題が残る。今ばらつきを求めたが，そのばらつきはあくまでも $\bar{x}$，すなわち，この標準溶液全体の値にばらつきをつけたに過ぎない。この標準物質を瓶に小分けし，売るのであれば，その瓶に入っている溶液の値にばらつきをつけなければいけない。

もう少し考えてみよう。$\bar{x}$ と値づけられた大量の標準物質を瓶に小分けすれば，ある瓶に入っている標準物質は，$\bar{x}$ から，瓶間の濃度差の分散だけ値が外れていることが期待されるであろう。つまり，全体の濃度につけられた不確かさに更に瓶間の濃度差の分散が丸々 1 個含まれるのである。よって，小瓶に入った標準物質のばらつきは，

$$S_{c2} = \sqrt{S_c^2 + \hat{\sigma}_A^2} = \sqrt{0.09922^2 + 0.04544} = 0.2351 \text{ mg/L} \quad \cdots\cdots(30)$$

となるのである。

よって，標準物質など，実際に測定したものに値をつけるのではなく，たくさん量があるものからいくつかサンプリングし，その量の値を求め，全体の値と考えるときには，その全体の値と小分けしたものの値では，ばらつきの構造が変わってくる。このようなことは特に破壊量の測定でよく起こる。破壊量とは，純粋な意味での繰り返し測定が不可能で，一度測ってしまうとその測定対象物が破壊されてしまい，二度と同じものを測定できない，というものである。コンクリートの強度試験や，金属の引っ張り試験，硬さ試験などがこれにあたる。また，一度瓶を開封し測定したものを売りには出せない，という意味では，標準物質も破壊量にあたる場合もあるだろう。

破壊量は分散分析法を用いないとばらつきが算出できないことが非常に多い。また，入念に実験の計画を練らなければ，求めたいものが求まらないことも多々ある。このような点に注意し，実験を行って欲しい。

6　複雑な分散分析

ここでは一元配置以外の複雑な分散分析を紹介し，一元配置との違いを簡単に解説する。

多元配置の分散分析は単に因子が増えるだけであるが，交互作用という項が因子の項以外に出現する。交互作用の詳しい話は割愛するが，ばらつきの推定に交互作用が取り扱いづらいばらつきをもたらすことがあり複雑になる。また，実験回数が非常に多くなることから 3 元配置くらいが限界だと考えた方がよい。

さらに，実験のランダム化が不可能な場合が存在する。よくある例では，日内変動と日間変動を評価する場合である。このとき実験の順番を考えても 1 日目に繰り返しを 5 回，2 日目に繰り返しを 5 回…と順番に行うことしかできない。時間を戻すことはできないので当然である。このようなときには「不完備型実験」と呼ばれ（ランダム化が行える場合には「完備型実験という」），通常の分散分析法ではなく，枝分かれ法，分割法といった手法が用いられる。測定データ処理であれば特に枝分かれ法が重要である。枝分かれ法に関しては構造もさほど複雑ではないので，ぜひ専門書等で理解していただきたい。

7　最 後 に

連載第 1 回に，実験計画の大切さについて解説したが，この点は熟練した技術者，研究者の方々でも非常に盲点になっているところである。私のところにもデータの評価方法を教えて欲しい，という依頼が多く来るが，ほとんどの場合は，「このようなデータを取ったのだけれども，どのように統計処理したらよいかを教えて欲しい。」というものである。このような場合，元々の実験計画が悪いため，いくら統計処理を施したところで，あまりよい結果が得られないことがある。つまり，データを取得する前に測定量，測定方法，測定手順を決定し，実験を計画し，データ処理方法を決めてから実験を行う必要がある。これが計量管理における一番の基礎部分である。

最初は実験計画の立て方にとまどうかもしれないが，慣れればそう難しいものではない。今後は，漠然と実験を始めるのではなく，入念な実験計画の構築から実験をスタートして欲しい。

また，今回は連載の最後であるので統計に関する参考文献をあげたいと思う。

まず，統計を学習する上での一番の基礎が解説されているものは，

1）田中秀幸，統計学　入門編・初級編，日本計量振興協会

である。筆者が執筆したものである。本連載の第 1, 2 回目の内容を平易に解説したものである。日本計量振興協会からの直販である。

次に，統計をある程度系統立てて学びたい方には，

2）東京大学教養部統計学教室，統計学入門，東京大学出版会

をはじめとする，大学の教養課程で用いられる教科書がよい。

分散分析について初歩から学びたい方は，

3）石川　馨，米山高範，分散分析法入門，日科技連

がよいが，絶版のようで手に入れるのが難しいが，図書館では蔵書に含まれているところは多いようだ。この本が他の分散分析の入門書より優れているのは，他の入門書では F 検定までしか解説しないものが多い中，分散の期待値の算出法が細かく解説されているところである。

データ解析のための統計を本気で学びたい方は，

4）近藤良夫・舟阪　渡，技術者のための統計的方法，共立出版

がよい。この本も絶版であるが，図書館の蔵書に含まれているところは多い。この本はデータ解析における統計の大部分が網羅されており，良著である。また，分散分析の構造に関する解説の中身も非常に濃い。

また，4）は絶版であるが，4）のエッセンスを抜き出したような本である，

5）安藤貞一・松村嘉高・二見良治，技術者のための統計的品質管理入門，共立出版

は現在でも手に入れることができる。

また，測定における統計に関しては JIS 規格もいろいろある。標準物質の値付けに関しては，

6）JIS Q0035：2008，標準物質−認証のための一般的及び統計的な原則

が重要である。今回解説した標準物質への値付けに関して，詳細に規定している。

また，標準物質に限らず，

7）JIS Z8402 シリーズ，測定方法及び測定結果の正確さ（真度及び精度）

8）JIS Z8402−2：2008，測定の不確かさ―第 2 部：測定の不確かさ評価における繰り返し測定及び枝分かれ実験の利用の指針

は主に分散分析法を用いたデータ処理法についての解説がなされている。

以上の本のみがお勧めというわけではないが，これらの本は統計を勉強する人たちにとって理解しやすい参考書であると思う。

田中秀幸（Hideyuki TANAKA）
産業技術総合研究所計測標準研究部門物性統計科応用統計研究室（〒305−8563 茨城県つくば市梅園 1−1−1　産総研中央第 3）。筑波大学大学院工学研究科修了。博士（工学）。≪現在の研究テーマ≫計測における不確かさについて。

新刊紹介

アミノ酸と生活習慣病
――最新アミノグラムで探る「いのち」の科学――

杤久保 修・安東敏彦 著

アミノ酸がタンパク質の構成成分でとても重要な栄養成分であるということは誰でも知っている。しかし，血液中のアミノ酸濃度やそのパターンと，生活習慣病（癌，メタボリックシンドローム，肝疾患，腎疾患）とが，これほど密接な関係にあるとは知らなかった。本書は，長年横浜市立大学医学部で循環器病や生活習慣病の専門家として活躍されてきた杤久保 修教授とアミノ酸研究で優れた業績のある味の素㈱ライフサイエンス研究所の安東敏彦主席研究員による，アミノ酸の血液中濃度に関する啓発書である。栄養失調と栄養過多，循環器病，肝臓病，腎臓病，消化器疾患，呼吸器疾患，皮膚疾患，ストレス，癌などで代謝物としてのアミノ酸がどのようにかかわり，その血液中の濃度がどのように変化するかが詳しく書かれている。本書の中では，アミノ酸パターンが車輪図（レーダーチャート）として表されていたが，疾病ごとに変化するアミノ酸が異なるので，結果として病態特有の凸凹を示す車輪図が出現する。病態をアミノ酸でビジュアル化できる点は特に興味深い。また，この膨大な研究データを得るためにはアミノ酸分析の超短時間化が必要である。そのための革新的技術進歩についても詳しく書かれている。現在，疾病ごとに様々なバイオマーカーが発見され，診断への応用研究が進められている。しかし，血液中のアミノ酸パターンで疾病の早期診断や病態把握ができる時代の近いことを予感させられる。

（ISBN 978−4−7895−5435−0・B 5 判・175 ページ・3,000 円＋税・2010 年刊・女子栄養大学出版部）

役にたつ イオンクロマト分析

㈳日本分析化学会
イオンクロマトグラフィー研究懇談会 編集

本書は，イオンクロマトグラフィー（IC）研究懇談会に所属している IC の基礎・応用分野の研究者および IC 関連装置を開発・販売している企業の研究・技術者によって執筆された実用書である。最新の IC 技術のハードウェアとソフトウェアを一体としてとらえ，IC 装置を使用している技術者・研究者に技術情報を提供することが重要とであるとの考えに基づいて出版された。構成は，第 1 章：IC の歴史・構成，第 2 章：IC の分離科学，第 3 章：IC の基礎技術，第 4 章：IC の基本操作，第 5 章：IC の公定分析法，第 6 章：IC の応用，Q&A, IC の将来展望などから成っており，IC に関することが網羅されている。どの章においても実際の測定例が測定条件とともに豊富に掲載されていてとても実用的である。Q&A では，様々な疑問や問題に対する解決法が示されており，入門者にもわかりやすい内容となっている。コラムとして書かれた「こぼれ話」も興味深く，巻末にまとめられている IC 用カラム一覧は，実際に使っている人にとってありがたい情報である。すでに IC を使っている人も，今後 IC を使う予定の人にとっても有用な情報が満載であり，必携の一冊と言える。

（ISBN 978−4−87211−973−2・B 5 判・228 ページ・3,400 円＋税・2009 年刊・みみずく舎）