６．調査結果の整理および解析

6.1 入力データのチェック

本調査では、多数の化学物質を多数の地点で測定するので、正確なデータをコンピューター上に間違いなく入力・保存することが、解析の第一歩として重要である。また、長期間にわたりデータの蓄積を行うため、過去にさかのぼってデータの訂正を行うことは困難である。したがって、間違いの可能性があるデータはなるべく早く発見し、適切な訂正を行うことが重要である。このためには、範囲のチェック、論理的にあり得ない入力でないかのチェックなどが有益である。

データの作成においては、分析試料の作成、化学分析および入力の各段階でエラー（主としてケアレス・ミス）の入り込む余地が多い。このようなエラーを検出するためには、入力後、少なくとも次のようなチェックが必要である。

- 範囲チェック：分析値は、上限が予想できるものが多いので、適当な範囲を設定することにより、桁違い等のエラーを発見する。
- 論理チェック：サンプリング日や分析日等であり得ない数字が入っていないかどうか。
- 関連チェック：入力項目間に矛盾がないかどうか。

これらのチェックを行うことにより、入力時のミスをできるだけ回避するとともに、できるだけ早い時期に、転記ミスやサンプルのとり違い等を発見できる可能性がある。

なお、ミスを発見した場合には分析者等に問い合わせ、数字の確認を行う必要がある。

また、物質濃度のみならず、ともに報告される一般項目や関連情報も、等しく重要なデータの要素であり、同様のチェックを行うべきである。

6.2 解析上の留意点

検出限界未満の測定値についても情報を最大限に生かせるよう留意する必要がある。不検出として測定値が報告されていない場合でも、それは、測定値がゼロから検出限界未満の範囲であるという情報を示している。検出限界未満で値の表示がある場合は、検出限界以上の値に比べ相対的に精度が低いと考えられるものの、最も確からしい値として計算に使用することが可能である。

特に相対誤差の大きな場合や、誤差の大きさにばらつきのある測定をまとめて評価する場合には、それらの誤差の影響が経年変化の傾向の判断に与える影響を含めて解析されることが望ましい。この場合、検出限界未満の測定値もあわせて、統一的な取り扱いがされることが望ましい。

場合によっては、一部の値の重みを小さくするような操作（例えば調整［trimming：最大値と最小値から一定の比率の順位にあるデータを取り除く］あるいはウインザライズ化［Winsorized sampling：最大値と最小値から一定の比率の順位にあるデータを、それ以外のデータの最大値と最小値で置き換える］など）により、外れ値と考えられる値の影響を小さくすることができる可能性がある。

多くのデータを処理する場合、仲間からとび離れた値が混入することがある。これを外れ値と呼ぶことにする。外れ値を生ずる原因としては、分析機器の読み違いや換算の際の桁違い、転記ミス等が考えられるが、これらの誤りについては追跡できる限りはその原因を追求すべきであり、原因が明らかになれば訂正可能である。外れ値を摘出することのさらに重要な意義は、異常値と判断した根拠、統計学的に言えば、データの構造に仮定したモデルが間違っていたために生じたかもしれないことを警告できることである。例えば、10 個のサンプルのうち、1 個だけ他と異なる値を示した場合、そのサンプルは海流等の影響で汚染の程度の異なる海域からのサンプルかもしれないということを警告してくれる。この場合はサンプルの代表する海域は異なるものとして取り扱う必要がある。

さまざまな要因による測定値の経年的なばらつきを取り除き、変動の大きな傾向に注目するために（加重）移動平均などの平滑化［smoothing］を行うことが有効な場合がある。変動の大きな傾向の直感的な把握や、大きな傾向に対する解析がより容易になる。また、傾向にさらに注目する場合には、移動平均値ではなく、移動中央値を用いることも一つの選択肢である。ただし、平滑化に際しては、原データの傾向を適切に反映した結果が得られているかどうか、注意が必要である。

本調査で得られるデータのように、分布形が正規であることが必ずしも期待できないか、測定量に対して意味のある変数変換によっては正規近似が困難である場合に、集団（例えば地域や調査年次など）ごとの特性の比較を行う方法の一つとして、順序統計量を利用することを考慮すべきである。順序統計量とはデータを大きさの順に小さい方から並べたときの値をいう。例えば 11 個のデータでは中央値は 6 番目のデータの値である。この値を一つ以上用いて位置の比較を行うことができる。順序統計量を用いる場合は、ND や TR があっても、ND<TR という関係が保証されれば、特に特定の数値を代入する必要はなく、そのままのかたちでの比較が可能になる。

なお、疎水性の化合物は有機物に吸着した存在形態が予想され、また同一地点であっても毎年度、均質な試料を採取することは困難であるので、底質中の有機物含量を指標値として用いて測定値を補正することが有効な可能性がある。

6.3 測定濃度の経年的な変動傾向の判断

経年的な変動傾向を評価しようとする場合、その目的と、また適用手法に応じて、対象とする期間や間隔が短かすぎることや長すぎることのないよう配慮する必要がある。

全期間において増加あるいは減少の傾向が認められるかどうかについては、例えば、年度に対する測定値の線型回帰係数が、統計的に有意に、ゼロでない正あるいは負の値を取るか、を判断の一助とすることができる。

通常の一次回帰モデルを用いる場合、傾きの値が統計的に有意にゼロでないことを言うためには、測定年が十分に多いか、年に対する線型な変化の傾向が十分に明瞭であるデータが得られている（回帰の決定係数 r^2 の値が大きい）必要がある。例えば、10 個の測定値がある場合には $p<0.05$ で傾きがゼロでないことを言うために、およそ決定係数 $r^2=0.40$ 程度以上の回帰が得られるような測定値が得られている必要がある。なお、回帰モデルによる有意性の検証とは別に、どの程度のばらつきのあるデータまでを経年的な変動傾向の判断対象とするのか、測定誤差に比して意味のある変動と言えるのか、などをあわせて検討する必要がある。

期間内に、減少から増加へ、あるいは増加から減少へ転じるピークが存在すると認められるかどうかについては、例えば、一次差分 $x'_t = x_t - x_{t-1}$ の変動を判断の一助とすることができる。この解析を行う場合には、平滑化を行うことで、大きな傾向をより適切に判断できる可能性がある。

全地点での全体的な傾向の判断が必要な場合、また、何らかの理由で、各地点での傾向を個別に議論することが適切でないと判断される場合には、経年的な変動傾向を全地点について総合的に評価する。一つの方法として、各年度における、全地点についての平均値あるいは中央値等の、経年的な変動傾向を評価することが考えられる。

6.4 データの総合的な解析

化学物質による環境の汚染は、個々の物質ごとに独立に発生することは少ない。また、対象物質濃度のみならず、さまざまな情報を合わせて解析することで、新たな知見が得られる場合がある。したがって、可能であれば、物質間の相関関係や関連した情報を考慮した総合的な解析を実施する。この際、さまざまな多変量解析法のうち、解析の目的に応じて適切な手法を選択して適用するべきである。また、このような解析を行う際には、解析に含める関連情報の精度についても十分考慮する必要がある。

以下に、多変量解析手法の一つである主成分分析について概略を述べる。主成分分析（Principal Component Analysis, PCA）は、互いに相関のある他種類の特性値の持つ情報を、情報の損失が最小になるように互いに相関のない少数個の総合特性値に要約するという方法である。また、そのことによりデータの解釈の助けとする。

すなわち p 個の特性値、（p 変量 ここでは、各検体ごとの化学物質の濃度）x_1, x_2, ..., x_p のもつ情報を、もともとの特性値の線型結合で表され、互いに無相関で、順番に減少する分散を持つ m 個（$m<p$）の総合特性値 z_1, z_2, ..., z_m—これを第 1,第 2, ..., 第 m 主成分（principal components）と呼ぶ—に要約する。

なお、主成分分析を適用する場合には、特に高い値や特に低い値の少数の測定値に結果が大きく影響されることを防ぐために、なるべく変数の分布が正規分布に近い方が好ましい。したがって、必要であれば対数変換等の変数変換を行う。

得られた結果は、化合物および調査地点のグループ分けやその特徴を議論するために供しうる。