化学物質の１－オクタノール／水分配係数と生物濃縮性との関連性

に関する統計学的考察に関する報告

平成１５年９月１６日

近畿大学農学部教授　米虫節夫

平成１５年５月２０日付け経済産業省化学物質安全室による資料－８「化学物質の魚介類の体内における濃縮度を判定するための知見の取り扱いについて」（以下、「資料－８」と略）中の「化学物質『ーオクタノール／水分配係数』を魚介類の体内における濃縮度を判定するための知見として取り扱うことについて（案）」(2 ～ 22 ページ)に記載されている当該課題について、報告書執筆者（以後、「報告者」と略）の過去の経験を基に以下のように判断したのでここに報告する。

１．結論

平成１５年５月２０日付け経済産業省化学物質安全室による資料－８「化学物質の魚介類の体内における濃縮度を判定するための知見の取り扱いについて」の４ページ「４．判断基準の見直し（案）」（３）に書かれた結論「現行の「３未満」を「４未満」に変更する」の推論は、概ね妥当と考える。この変更を用いた時の BCF 値の中心値の推定値は 106.7 倍であり、95%信頼区間の上限値は 2307 倍である。

しかし、BCF の 1000 倍未満を強調したいのであれば、「**3.5 未満**」の方がさらによい。この変更を用いた時の BCF 値の中心値の推定値は 47.4 倍であり、95%信頼区間の上限値は 1023 倍である。

２．推測統計学的方法を適用する前提条件などの検討

2.1.母集団規定とランダムサンプルの問題

数理統計学的方法（推測統計学的方法、略して「推計学的方法」と云うことが多い）を用いた推論を行うときには、規定した母集団から適切な方法でランダムサンプルが得られていると仮定して行われる。母集団から無作為に得られたランダムサンプルを用いたときのみ、推計学的方法は正しく適用できる。自己に都合の良いサンプルのみを集め解析しても、正しい結論は得られないのである。ゆえに、取り扱うデータが推計学的検討に絶えうるものであるかどうかの検討は、第一義的に重要である。

さらに、得られた結論は、規定した「母集団」に対してもみ適用される。ゆえに、母集団規定の検討が必要になる。

2.1.1.母集団規定

本解析における母集団は、「魚介類を用いた濃縮度試験の結果において一定の濃縮制が認められた化学物質」＜母集団規定１＞（資料－８、p.2、2(1)１行目)とまず規定されているが、これには限定条件が付いている。すなわち、「得られたＢＣＦ最大値５００倍以上～平均値 10,000 倍以下」の６４物質、「ＢＣＦ最大値５００倍未満」の１４３物質、合わせて２０７物質（資料－８、p.2、2(4)１～ 2 行目)とされているので、母集団規定１のうち、ＢＣＦ平均値 10,000 倍以下の物質と云うことになる＜母集団規定２＞。

さらに、「得られたＢＣＦ最大値５００倍以上～平均値 10,000 倍以下」の１１１化学物質のうち、「logPow の測定値のないもの又は logPow 測定値が使用できないもの」＜母集団規定３＞４７物質が除外されているので（その結果、111 - 47 = 64)、この条件に該当するものは母集団に含まれない。logPow の測定値を基にして、BCF 値の検討を行おうとしてい

るのであるから、logPow の測定値のないもの又は logPow 測定値が使用できないものは、当然利用することは出来ず、検討される母集団の中にも入ることは出来ないので、この除外は問題が無かろう。

BCF 値が 500 倍未満のモノと、500 ～ 10,000 倍までのモノとの２グループのデータが合併されているが、これはもともと連続値である BCF 値を 500 未満（過去において、高濃縮性と判断されていないモノ）とそれ以上というように２分割したものであるから、この合併は母集団規定の点で問題にならない

ゆえに、本解析結果の適用される母集団は、母集団規定１に示された物質のうち母集団規定２に該当し、母集団規定３を除外した物資と云うことになる。

2.1.2. ランダムサンプル

規定された母集団から乱数を用いて得られたデータが、ランダムサンプルであるが、本報告書のような解析においては、規定された母集団に属する全てのデータを用いている。手持ちのデータ、すなわち現在得られているデータを全て用いて解析したということであるが、現在得られているサンプルは、今後合成されたり発見されたりする物質も含めた「無限母集団」から、ランダムに合成されたり発見された物質であると考える。このように考えると今解析対象としようとしている物質は、無限個の母集団規定に合致する物質中から得られたランダムサンプルとして、取り扱いすることが出来る。

ゆえに、手持ちのデータは、規定された母集団からのランダムサンプルとして取り扱っても良いことになる。

2.2. 相関・回帰分析に関する問題点の検討

本報告書においては、相関分析と直線回帰分析が用いられている。まず、この解析方法の妥当性について検討する。

2.2.1. 相関分析利用の問題点

n 個のデータセット(x_i, y_i)、i=1, 2, ・・・n を用いて pearson の相関分析（通常のテキストで学ぶ相関分析で、両変数間に直線的な関係の存在を仮定している）を行うには、変数 x と変数 y との間に、二次元正規分布が仮定されていなければならない。しかしながら、本解析に用いているデータは、x に該当する変数が logPow、y が logBCF で、両者共に測定値であるためそれぞれ一定の誤差を持つ値となる。しかし、x の平均値と y との平均値で決定されるある一点を中心とする二次元正規分布にしたがっているとは思われない。ゆえに、このデータの解析に相関分析を適用するのは適切ではない。

資料－８、p3、「３．結果及び考察」(1)において相関関係を求めている部分は、用いているデータの性質からみて、正しい解析とはいえない。

しかし、相関分析においては通常、相関係数 r を用いた推論を行うのであるが、資料－８の上述部分では、「決定係数 R^2」を用いた考察をしている。この決定係数は、回帰分析における「寄与率」と同じものであり、資料－８の文章では相関分析を行っているが、その目的は回帰分析の前段階的な扱いになっており、首尾一貫した記載になっていない。

2.2.2. 直線回帰分析の前提条件の検討

2.2.2.1.　データの前提条件

独立変数 x が決定された条件下で、従属変数 y の値を誤差を考えて推測するための手法、それが回帰分析である。n 個のデータセット(x_i, y_i)、i=1, 2, ・・・n を用いた直線回帰分析を行うためには、変数 x には誤差が無く、変数 y のみが正規誤差を持つことを仮定している。

先に書いたように x に該当する独立変数が logPow、従属変数 y に該当する変数が logBCF であり、両者共に測定値で誤差を持つ。ゆえに、この種の直線回帰においては、得られた logPow 値には、誤差はなく、logBCF にのみ誤差を持つと仮定して解析されており、数学的な仮定からは外れているが、生物分野の解析においては一般的に利用されている解析手法である。また、現実的には、得られた誤差を持つ logPow 値から、推計学的に最も確からしいＢＣＦ値を推測する手法としては、有力な解析・推定手法である。

2.2.2.2. 直線回帰の妥当性

直線回帰を採用することの妥当性については、n 次曲線による曲線回帰モデルを仮定した分散分析を行うことより検定することができるが、図１からみて、明らかに単調増加の様相を示しており、敢えて直線回帰の検定は不必要であり、直線回帰を用いることに問題はない。

また、Pearson の直線回帰を前提とした相関分析で得られている相関係数や決定係数の値の検定・推定からも、直線回帰の妥当性は推察される。

2.2.2.3. 回帰直線による検討範囲の問題

回帰分析における推定の範囲については、変数 x である logPow の測定値の範囲０以上６以下の範囲内でのみ論議されている。

数理統計学で用いている回帰直線は、最小二乗法による単なる曲線（この場合は直線）の当てはめにすぎない。ゆえに、得られた回帰式を独立変数 x の範囲外へ外挿して理論を展開する時には十分な検討が必要となる。しかし、本解析ではその様な外挿はなされておらず、回帰分析の利用において検討範囲については問題がない。

３．資料－８、3 ～ 4 ページの記載事項の検討

資料－８の中心部分である記載事項について検討する。ここで行った推測統計学的検討に於る計算は、JUSE-Stat Works Ver. IV 3.0plus を用いて行った。

3.1. 「３．結果と考察」に対する検討

(1) 相関分析については、2.1.で検討した。

(2) 得られたデータに関する事実の記載であり、推計学的考察は含まれていない。

(3) 図１に示された回帰直線、決定係数などの値は、報告者が検算した値と有効数字３桁において一致している。分散分析の結果、誤差分散の大きさは 0.455 で、推定の誤差を示すその平方根は σ = 0.675 となる。回帰直線から± 1 の範囲は、1 ／ 0.675 ＝ 1.482 倍になり、

正規分布表から見て、約 86%という数値が計算としては得られる。ゆえに、この「約 86%」という数値は正しい。

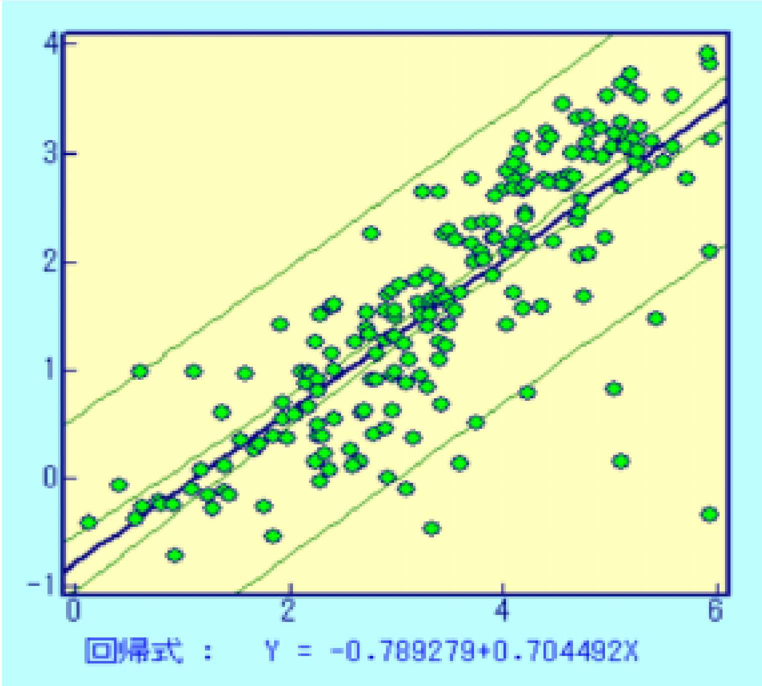
しかしながら回帰直線の両側に± 1 又は± 1.0 σの幅を引いて回帰のばらつきについて検討する方法は、簡易法としては良く用いられるが、正確な検討方法としては、正しくない。回帰分析においては、変数 x の平均値 3.44 の所が一番推定精度が良く、上述の約 86%という数値が正しいのは、この x の平均値の部分においてのみである。一方、x の平均値から x の数値が大なる方向又は小なる方向に外れるに連れて二次関数的に従属変数 y の推定精度は悪くなる。

(4) 上述のごとく、回帰直線における 95%信頼区間は、変数 x のところで最小の幅になり、独立変数 x の平均値から大小の両方向に外れるに従い二次関数的に幅が広くなる。図３では、そのことを無視して、変数 x の平均値の部分で得られる 95%信頼幅を他の部分にまで適用し、得られた回帰直線と平行な 2 直線でもって 95%信頼区間と称しているが、これは正しい信頼区間の求め方ではない。

さらに、「95%信頼区間：± 1.346」とあるが、この値は、(3)で示した誤差分散の平方根σ= 0.675 の 2 倍の値になっている。95%信頼区間のためには厳密には、± 2 倍ではなく、± 1.96 倍が正しい x の平均値 3.44 の位置における 95%信頼区間のはずなので、この幅は少し広い目になっている。

二次関数で示される直線回帰の 95%信頼区間の上限値、予測値及び下限値を、正しい推計学的方法で計算すると次のようになる。

logPow	データの９５%信頼区間上限	予測値	データの９５%信頼区間下限
0.5	0.913	-0.437	-1.787
1.0	1.260	-0.085	-1.430
1.5	1.608	0.267	-1.073
2.0	1.957	0.620	-0.718
2.5	2.307	0.972	-0.364
3.0	2.658	1.324	-0.010
3.5	**3.010**	**1.676**	**0.342**
4.0	**3.363**	**2.028**	**0.694**
4.5	3.716	2.381	1.045
5.0	4.071	2.733	1.396



「95%信頼区間から外れる物質は、１物質をのぞいて、・・・」との記載は、上記の計算結果から見ても同じ結果となる。物質 613A は、logPow = 0.6 で、logBCF3av = 0.99 であるが、正しい 95%信頼区間の上限の計算結果は 0.982 であり、わずかに外れている。

計算値は、若干誤っているが、結果は、同じになった。

(5) 図３の回帰直線より右下方向に外れる物質が、「安全サイド」のモノとして、それに属する８物質を除外し再度回帰分析を行っているが、報告者はこの場合、その必要はないと考える。

臨床データの解析などで、治験に組み込んだ患者をよく調べてみると、当初の母集団規定に合致していないことがわかり、解析集団から除外して再度解析をすることは良くあることである。このような一部のデータを除外した解析を行うことにより、より精度の高い推測が可能となる。しかし、純粋数学の学者などはこの種のデータの除外を極端に嫌っている。報告者は、主として現場的データの解析を主として行ってきたので、母集団規定に違反するデータの除去にはやぶさかではないが、本解析のようなときに 「安全サイド」のデータを除外して再解析するのは、あまり適切な方法とはいえない。もし、除外するのであれば、それに見合った母集団規定を考えねばならないからである。

ゆえに、検算した結果、計算結果などには誤りはないが、報告者は、この解析は不必要と考える。

(6)　(5)において除外したデータセットによる解析結果が、図４に示され、それを用いた推論がなされている。しかし、前述のように報告者はそのような除外を行わず、図３を基礎とした推論を行う方が適切と考える。その立場で、以下に考察する。

回帰分析の結果から計算されるＢＣＦの中心値並びに95%信頼区間の上限値に着目する。logPow の値が、3.5 および 4.0 に対する logBCF3av の中心値および 95%信頼区間の上限値は、それぞれ 1.676 と 3.010 および 2.028 と 3.363 である。これを元のＢＣＦ値に戻すと、47.4 倍と 1023 倍および 106.7 倍と 2307 倍となる。

logPow	logBCF3av	BCF3av	
3.5	1.676	47.4	中心値
	3.010	1023	上限値
4.0	2.028	106.7	中心値
	3.363	2307	上限値

3.2. 結論：判定基準の見直しに対する考察

3.2.1. 論点のまとめ

高濃縮性を判定するとき、BCF 値が 5000 倍以上かどうかで判断していると(1)に記載されている。また、資料－８の文中には、この値が 1000 倍未満であるかどうかが随所に書かれている。故に、5000 倍以上は高濃縮性の判定、500 倍以下は従来からも高濃縮性とは判断している。この関係は、次のように示される。

BCF の値	判定
500 倍未満	高濃縮性ではないと判定する
500 ～ 5000	従来は、高濃縮性ではないとの判断をしていたが、 この範囲内に両者を区別する限界値を設定したい。
5000 倍以上	高濃縮性と判定する

すなわち、500 ～ 5000 の範囲内のどこかに「高濃縮性ではない」と判断できる限界値を設定したい。出来ればこの「500 倍未満」という限界値を「1000 倍未満」にしたいというのがこの資料－８の趣旨のように思われる。

資料－８の作成者は、この目的で logPow 値の「4.0 未満」を採用しているが、前述のような計算方法の違いで、「4.0 未満」を 1496 倍と計算し、「1000 倍未満」と論じている。

3.2.2. 結論

報告者の計算結果では、logPow 値が、4.0 に該当する濃縮倍数の 95%信頼区間の上限値は 2307 倍であり、1000 倍よりは少し大きい。一方、logPow = 3.5 におけるその値は 1023 倍となり、ほぼ 1000 倍である。

ゆえに、報告者は、得られた解析結果から、次のように結論したい。

BCF 値の 95%信頼区間の上限値「1000 倍未満」に重点を置けば、logPow の値は、「3.5 未満」が最も良いであろう。それほど厳密に、1000 倍に拘らず、5000 の 1/2 の数値である 2500 倍程度までならまだ良いであろうとの認識を示すならば、logPow の値は、「4.0 未満」が良いであろう。

４．報告者のスタンス

先に書いたように報告者は、数理統計学の専門家ではないが、数理統計学を現場に適応させる仕事に 40 年ほど関係してきた。

大阪大学大学院工学研究科醗酵工学専攻に在学中、研究テーマとして与えられた「微生物熱死滅の動力学的研究」を遂行する過程で、平板上に形成される生残菌数の処理を適切な統計処理の方法を駆使しておこなった。

数理統計学の基礎は、工学部及び工学研究科における授業により学ぶと共に、（財）日本科学技術連盟主催の品質管理 BC コース、実験計画法コース、オペレーションリサーチコース、臨床統計学コースなどでも学び、それら幾つかのコースの講師をすることにより、理論・応用両面から多くの数理統計学者らの指導を受けた。

大阪大学薬学部在職中も、微生物の死滅に関する研究を行うと共に、医薬品の微生物汚染に関する推計学的研究、無菌試験における試験サンプル数の数理統計学的考察、薬理学分野における「薬物の用量－反応関係」に関する著作の翻訳、新薬開発や医療情報の推計学的研究などに取り組んだ。

近畿大学農学部に移籍後も、製薬企業や臨床医師などの要請を受け新薬開発や医療情報の推計学的研究などに引き続き取り組んできた。ゆえに、この分野における論文は多い。

以上の経歴と経験から明らかなように、報告者は与えられたデータセットを用い、理論的に使用可能な推計学的方法を駆使して、可能な限り色々の結論を得ることを目的に、解析の仕事をしてきた。この報告書には与えられた情報を基に、報告者が通常行っているデータ採用方法、解析方法を用いて、検討可能な理論的検討と若干の解析を行い、妥当と思われる所見を記載した。

以上