

やさしい統計のお話 (基礎的な考え方について)

(クォーター-No.5 より)

私達は日常業務の中で、統計的手法を用いることが時々ありますが、「さてその依つて来たる所以は」と聞かれると一寸説明にとまどうこともあるのではないのでしょうか。過日、お花見の宴で同席した知人が、統計的なものの考え方についてエピソードを交えて判りやすく話していたので、早速職業意識?を働かせて、クォーターへの投稿をお願いしました。さらに今後も引続いて書いて下さるとのことですのでご期待下さい。

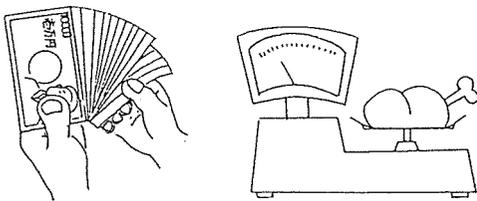
1. はじめに

私達は、サンプリングをし、測定をしてデータを取りそのデータを見て何らかの処置を取る、という仕事に携わっています。従ってデータをどう扱うか、どう解釈するかは大変大事なことです。技術的知識と判断が基礎になることは当然ですが、データの統計的判断もまた必要になります。

統計の講義といえますと、先ず確率、統計値、そして平均値、標準偏差の計算ということになりますが、その前に是非理解して置いて頂きたいことがあります。一番基礎的な考え方なのに忘れられていることも多いようです。とかくテクニックすなわち定義や計算方法に飛び付きがちですが、ここではその基礎的な考え方についてお話をしたいと思います。

2. データを取るのには目的がある。

健康診断で身長と体重を測りますが、その目的はその人の健康状態を見る一つの目安を知るためです。預金するため銀行へ行き、1万円札28枚と千円札20枚を出しますと、窓口の人は2、3回お金を慎重に数えます。合計30万円あるかどうか確認するためです。このような単純な測定の例でも必ず目的があります。そしてこれらの場合は誤差(「真の値との差」とだけ説明しておきます。)は問題になりません。お金を扱う社会では誤差を認めないそうです。千円や壱万円違ってもいいや、というわけにはいきませんから当然でしょう。もう一つ大事なことは、後で詳しく述べる母集団という概念も必要ないということです。



少し問題のある測定の例をあげましょう。スーパーで100g450円の牛肉を300g位買うとします。店のハカリで測ったら316gと出て、 $450 \times 3.16 = 1422$ 円支払います。このハカリは正しいでしょうか。タクシーに乗って1110円の料金が出ました。このメーターは正しいでしょうか。これらの場合には誤差が問題になりますが、取り引きに使われる計量器は、法律により、県の計量検定所が検定や検査を定期的きちんとして、認められる誤差内にあることを保証して呉れていますから安心です。さてこの場合も母集団ということ、取りあえず考えなくても良いでしょう。

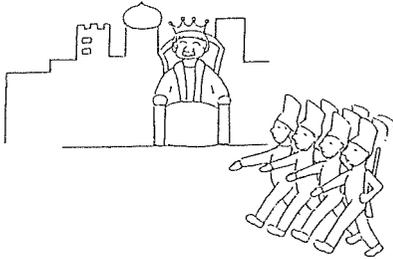
3. 母集団という考え方

ある自動測定器の保守点検の仕事を考えましょう。きちんと正常に作動しているかどうか、を確認する目的で定期的に巡回して何らかのデータを取り、それにより判断します。知りたいのは、その測定器の、その時点での「機能的状況」なのです。この状況が母集団という考え方です。あるいは、ある工場の排水をサンプリングして原子吸光で微量金属成分を分析することを考えますと、測定したデータから、その工場の排水の状況が基準に合っているか、異常が無いかを知るのが目的ですから、この場合は「その時点でのその工場の排水の状況」が母集団となります。またそのときの分析方法も一つの母集団となります。同じ試料を分析しても測定の度にデータが少しずつ違うのが一般ですから、「この分析法の本来の実力」(正確さとバラッキ)という母集団を考えなければなりません。

4. 母集団と人口は同義語か

母集団という言葉は完全に統計用語です。英語ではPopulationと言います。Populationを辞書で引くと「人口、住民数」が一般訳ですが、実は母集団と人口は同じ意味であった、という面白い話があります。

何世紀か前ヨーロッパの専制王国の王様は、戦争に備えて何人徴兵出来るか、税金をどれだけ取り立てることが出来るかの基礎資料として総人口つまり母集団を知る必要があったわけですから、人口と母集団は同義語だという説です。ついでにいいますと、統計の英語はStatisticsです。国や州のことをStateといいま



から、もともと統計というのは、「国を治めるための資料」の意味だったようです。現在でも日本では総理府統計局が中心になって国家行政に必要な各種統計が取られ、公表されています。何年かに一回、莫大な人手とお金をかけて国勢調査をやります。いわゆる全数検査で、日本中の人口の母集団がすっかりわかったということになります。極めて重要な資料となります。然し本当に正確かどうかについては若干の疑問が残ります。調査ミスも大分ありそうですし、日本中では片時も途切れなく、人間が生まれたり死んだりしています。調査ミスは一応置いておくとしても、ある時点でのデータ、すなわち世の中の流れを止めて測定したもので、いわば「静的」な姿を示しています。

国勢調査のデータにケチをつけているのではありません。貴重なものですし、極めて役立つものと思います。言いたいのは、私達の技術分野ではそのような全数検査は不可能なことが多いし、また必要が無いことが多いということです。

5. サンプリングと母集団の関係

ある自動測定器の保守点検で、標準ガスや標準液でデータを取るはその測定器の作動状態が正常かどうか、すなわち母集団の状況を把握することが目的である、と述べました。私達が扱う仕事でのサンプリングと測定は殆どこの考え方が適用出来ます。サンプリングとは、対象とするもの全部を調べるのではなく、比較的少数の見本、試料を母集団の代表として採用する、という意味です。どのようにして試料を取るか、 n の数をど

うするかの問題はありますが、事実それしか出来ないケースが大部分です。先に説明した人口統計の例では、母集団である日本の総人口一億二千万人、いくら多いと言っても有限です。人手と費用を惜しまなければ、サンプリングでは無く、全数を調べることは出来ますこれに対して、自動測定器の母集団は動的であり、有限では無く、いわゆる無限母集団です。別の例で言いますと、ある電気部品を作る製造工程で、1日1回 $n=5$ でサンプリングしてある特性値を測定し、その工程が正常に稼動しているかどうかをチェックしたい、という場合母集団はその工程ですが、その工程で出来る部品は無限です。またある分析方法について、「標準操作法」を守ってきちんとやっている筈ですが、「標準試料」を使って週に1回 $n=2$ でチェックする、という管理をやっているとしますと、「その分析法の実力」が母集団であり、これも無限母集団です。無限母集団を対象とする限り、サンプリングしたデータから元の母集団の姿を推定しようというのが基本的態度です。統計学というと、全数を調べる、静的というイメージが浮かぶので、むしろ推計学、あるいは推測統計学という言葉を意識して使う人が多いようです。

6. 誤差を認知する

もう一つ基本的な考え方として「誤差」があります。誤差の性質や解析法についてはいろいろの参考書（例えば神環協ハンドブック「分析の誤差管理」）がありますのでここではこれ以上触れません。ただ基本的なことだけを言いますと、私達の測定には必ず誤差が付きまといま。従ってこの誤差を堂々と認知して、その規則性に注目して定量的に掴み、管理改善に積極的に活用して行こう、ということです。ただし認知したからといって大いばりではありません。より減らす努力が必要であることは勿論です。

7. むすび

以上統計の基礎的概念を述べて来ました。各種のテクニックについては数多くの本が出ていますので、それらで勉強してください。また電卓やパソコンで相当大量で複雑なデータも簡単に処理出来るようになりました。ただ強調しておきたいことは、データを見るときは、是非とも「このデータの元の母集団は？」「このデータの誤差はどのくらい？」の2点を頭に浮かべようにして下さい、ということです。