

データ評価のための統計的方法

——測定と統計の基礎知識——

1 はじめに

測定とは、ある物理現象をより良く知るために行うものであるが、測定したデータをどう解釈するかということは案外難しい問題である。データを解釈する際に変大有効であるのは統計的手法であり、適切な統計的手法を取得したデータに適用するとデータの解釈が非常に楽になるだけではなく、データ、グラフを眺めているだけでは見えてこない隠された性質までも明示することができるようになる。ただし、正しく取得されたデータに正しい統計的手法を当てはめなければ判断を間違う結果が得られてしまう。これを避けるには、測定についての十分な知識、統計についての十分な知識のほかに、測定と統計を結びつけるための知識が非常に重要になる。

本表題のシリーズでは統計的手法だけではなく、測定と統計を結びつけるための知識についても解説できればと考えている。シリーズは全3回を予定し、第1回の本稿は、測定と統計の基礎知識について解説する。また第2回では推定・検定について、第3回では分散分析法について解説する予定である。

本解説で用いられる記号については表1に凡例を示すので参考にして欲しい。

表1 凡例

母平均	μ
母分散	σ^2
母標準偏差	σ
標本平均	\bar{x}
標本分散	s^2
標本標準偏差	s
確率変数 x の期待値	$E(x)$
確率変数 x の分散	$\text{var}(x)$
確率変数 x の確率密度関数	$p(x)$
事象 A が起こる確率	$P(A)$
確率変数 x が区間 $a < x \leq b$ に含まれる確率	$P_{a < x \leq b}(x)$

Fundamental Knowledge for Reliable Analysis—Statistical Methods for Data Analysis: Part 1.

2 測定と統計

「測定」は何を知りたいがために行っているのでしょうか？ 例えば、ある溶液の濃度測定を10回繰り返して行い、その得られた測定値の平均値を算出した。このとき知りたい情報とは何だろうか？ 測定値の平均値を知りたいがためにこのような測定をしたのであろうか？ そうではない。ここで知りたいのは単に10回のデータの平均値が知りたいわけではなく、その溶液の本当の濃度を知りたいのである。しかし、測定値はばらつきを持つ。本当の濃度を完全に知るためには無限回の測定（またはそれに準じる回数回の測定）を行わなければならない。しかし、有限の時間に無限回の測定を行うことは不可能である。またはそれに準じる有限回の測定であっても時間・コストの面から難しい。よって10回の繰り返し測定のデータから、その溶液の本当の濃度を推定しているのである。

この例での本当に知りたい値とは、測定を無限回行ったときに得られるデータの集まりの平均値である。この測定を無限回行ったときに得られるデータの集まりのことを母集団と言い、その母集団の平均値のことを母平均と言う。この母平均をはじめとした母集団の性質を表す値のことを母数と言う。またその母集団からサンプリングされたデータの集まりを標本と言い、標本の平均値を標本平均という。この関係を図示したものを図1に示す。

つまり、何を測定するのか、その測定される量の定義、測定方法、測定手順が定まったときに（我々には知ることができないが）ある母集団が決定する。そして測定を1回行うことによって、その母集団からデータが一つサンプリングされて測定結果を得る。つまり測定とは、その測定の母集団からのサンプリングであると言える。

何回か測定を繰り返すことによっていくつかの標本を得、その標本を用いて母集団の性質を推定する。これが統計的視点から見た測定の本質である。

先ほど、母集団はその測定される量の定義、測定方法、測定手順が定まったときに決定すると言ったが、こ

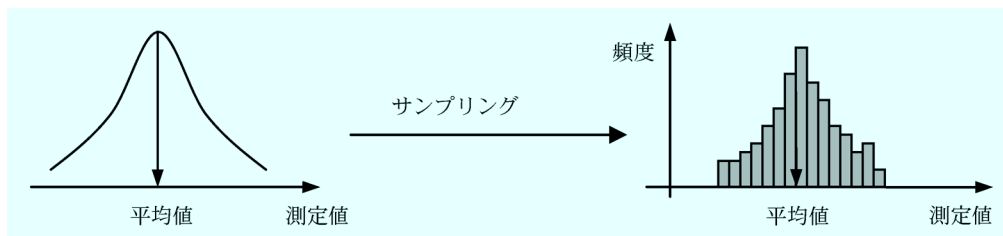


図 1 測定の母集団（左）と標本（右）

それは逆についても言える。つまり、測定を行うときには、量の定義、測定方法、測定手順を完全に決めておかなければならない。測定を行うごとに量の定義、測定方法、測定手順が変化したとすると、測定の母集団が測定ごとに変化し、測定されたデータは同じ母集団からサンプリングされたものとは見なすことはできない。また測定方法、測定手順が妥当なものでなければ、推定したいと思っている母集団とは異なる母集団からの標本によって母平均を推定してしまうこともある。よって、そのような標本から推定された母平均は信頼することができない。このことは当然のようで非常に見過ごされがちなことである。

例えば次のような実験を考えよう。製造装置を選定するために A 社、B 社製の製造装置を用いて同じ製品を製造し、その製品を比較したい。また、その製品を製造するには半日かかるとする。このとき何も考えずに実験を行うと表 2 で示すように行うことが多いだろう。

もしこの実験を月曜日の午前中から行ったとしよう。そうすると、製品を製造するのに半日かかるので、午前と午後で 1 日に 2 個の製品が製造できる。このことを考慮し、表 3 にいつ製品が製造されるかを示した。

表 3 を見ると、A 社の装置はすべて午前中に製造し、B 社の装置はすべて午後である。このようなデータを取ってしまうと、A 社製の製造装置で製造した製品と B 社製のものが何か異なる結果が出たとしてもそれは本当に A 社製、B 社製の製造装置が原因であるのか、それとも午前、午後に製造したことが原因であるのかの区別ができない。何も考慮せず実験・測定を行うとこのようなことが多発する。これを避けるためには「実験のランダム化」を行う。実験のランダム化とは、実験を行う順番を乱数表などを使って、ランダムに行うことである。実験のランダム化を行った実験の順番を表 4 に示す。

表 4 のように実験のランダム化を行えば午前・午後の要因が A 社製、B 社製の両方にほぼ均等に入ることが期待でき、純粋に製造装置の比較ができるようになる。

このように測定量の定義、測定方法、測定手順は質のよいデータを取得するための最重要項目であり、測定を行うときにはこの三つの項目を測定を行う前に完全に決定する必要がある。つまり、測定によって何を知りたい

表 2 実験の順番

製造装置\回数	1	2	3	4
A 社製	(1 回目)	(3 回目)	(5 回目)	(7 回目)
B 社製	(2 回目)	(4 回目)	(6 回目)	(8 回目)

表 3 実験の順番による不具合

製造装置\回数	1	2	3	4
A 社製	(月曜午前)	(火曜午前)	(水曜午前)	(木曜午前)
B 社製	(月曜午後)	(火曜午後)	(水曜午後)	(木曜午後)

表 4 ランダム化された実験の順番

製造装置\回数	1	2	3	4
A 社製	(月曜午前)	(月曜午後)	(火曜午後)	(木曜午前)
B 社製	(火曜午前)	(水曜午前)	(水曜午後)	(木曜午後)

のか、その知りたいことはどのような物理的特性を測定すれば達成できるのか、その物理的特性の定義をどのように実現するのか、その測定の手順をどのようにするのか、測定結果にどのような統計的手法を適用するのか、統計的手法を適用し得られた結果をどのように活用するのか、ということすべて決定した後に測定を行わなければならない。測定を行いデータを取得した後にそのデータに適用する統計的手法を考えるということを行うと、本当に知りたい情報はその測定結果からは得られないということがたびたび起こる。

3 分散と標準偏差

通常繰り返し測定を行った場合には測定結果として平均値を用いるが、その測定結果のばらつきの大きさも重要な情報である。適当な前提条件を置き、算出されたばらつきの情報を用いれば母平均の存在区間を算出することもできる。本章では標本のばらつきについて考える。

標本のばらつきを表すためには標本分散または標本標準偏差が用いられる。標本 x_i ($i = 1, \dots, n$) の標本分散 $s^2(x)$ は、

$$s^2(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \dots\dots\dots (1)$$

で表される。ここで、 \bar{x} は標本の平均値つまり標本平均を表す。また標本標準偏差 $s(x)$ は標本分散の平方根である。

式(1)より、標本分散とは各標本と標本平均の差の二乗和を $n-1$ で割ったものである。これは各標本と標本平均の距離（これを偏差という）の二乗平均を表している。本来であれば平均を算出するので標本数 n で割るはずであるが、標本分散を算出するときには $n-1$ の自由度と呼ばれるもので割ることになっている。これについては第6章で詳しく解説する。また分散は偏差の二乗平均であるので、例えば測定データの単位が $[g]$ であったとすると、分散の単位は $[g^2]$ となる。つまり、標準偏差では単位を元の測定量の単位に戻すために平方根を取るのである。

これを見てわかるように、正確な言い方ではないが標準偏差はデータの平均的なばらつきを表している。

統計ではばらつきは通常、分散のまま計算が行われ、最終的にばらつきを報告するときに標準偏差に変換される。この理由も第6章で解説する。

4 期待値と母分散

4.1 期待値について

第3章では標本分散、標本標準偏差について考えたが、本章では、標本だけではなく、母集団と標本両方について考え、標本と母集団の関係を解説する。

母集団と標本との関係を考えるに当たって最も重要なのは期待値である。期待値とは簡単に言うと「理想的にはこの値になる」という値のことである。

例えばサイコロを考えよう。サイコロを無限回振ったときの平均値はいくつになるだろうか？ サイコロは1から6までの面を持ち、また各面が $1/6$ の確率で現れる。よって、平均値は1から6までの面が同じ確率で現れるので3.5となるだろう。

つまり、このサイコロを振ったとき1から6のまでのどの目が出るかはわからないが平均的には3.5くらいが期待される、ということである。この3.5のことを期待値と言う。

これを数式で表すことを考えよう。「1が出るときの確率は $1/6$ 」というものを数式で表すと、

$$P(1) = \frac{1}{6} \dots\dots\dots (2)$$

となる。また、サイコロはすべての目の出る確率は等しく $1/6$ であるので、

$$P(1) = P(2) = \dots = P(6) = \frac{1}{6} \dots\dots\dots (3)$$

となる。この $P(A)$ というのは「 A という事象の起こる確率が $P(A)$ である」ということを表している。

次に期待値の算出は、

$$\begin{aligned} & 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} \\ & + 6 \times \frac{1}{6} = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} \\ & = \frac{21}{6} = 3.5 \dots\dots\dots (4) \end{aligned}$$

という式で考えることができる。これを一般式で表すと、

$$\begin{aligned} E(x) &= x_1 P(x_1) + x_2 P(x_2) + \dots + x_n P(x_n) \\ E(x) &= \sum_{i=1}^n x_i P(x_i) \dots\dots\dots (5) \end{aligned}$$

となる。 $E(x)$ は「 x の期待値」を表す。また、このときの x のことを確率変数と呼ぶ。確率変数とは値が確率的に決定する変数のことである。

また、サイコロの目は離散的な値であるので、期待値は式(5)となるが、測定値などの連続的な値の場合、期待値は、

$$E(x) = \int_{-\infty}^{\infty} x \cdot p(x) dx \dots\dots\dots (6)$$

と表すことができる。ここで、 $p(x)$ は確率密度関数といい、

$$P_{a < x \leq b}(x) = \int_a^b p(x) dx \dots\dots\dots (7)$$

によって、 x が a から b の間に含まれる確率を計算することができる関数である。つまり、連続分布のときには x がある値となる確率は0となる。なぜなら x は無限個のデータから構成されるため、無限個のデータの中からある一つのデータがサンプリングされる確率は0となるからである。よって、連続分布では x がある値からある値の間に含まれる確率というもののしか意味を持たない。

ここで、サイコロの期待値をもう一度見てみると、サイコロの期待値は3.5であるが、これは母平均と等しい。これを一般的に言うと、測定値 x があり、その母平均を μ とすると、

$$E(x) = \mu \dots\dots\dots (8)$$

が成立する。つまり、測定値は理想的には母平均になるということを表している。ここで、母集団についての性質を表す変数は通常ギリシャ文字が用いられる。母平均は通常 μ で表される。

このほかの期待値の性質を見てみよう。 c を定数とすると、

$$E(c) = c \quad \dots\dots\dots (9)$$

$$E(cx) = cE(x) \quad \dots\dots\dots (10)$$

が成立する。式(9) は自明であろう。言ってみるとサイコロの面すべてが3であるサイコロはいくら振っても3しか出ないということである。また式(10) に関して、サイコロの目が2, 4, 6, 8, 10, 12の目があつたとすると、期待値は7になることはすぐにわかる。

次に x, y がともに確率変数であるとき、

$$E(x \pm y) = E(x) \pm E(y) \quad \dots\dots\dots (11)$$

が成立する。これもここでは詳しく解説しないが、二つのサイコロを用意して出た目の和の期待値がどうなるかを考えれば理解できるであろう。

最後に確率変数 x, y が互いに独立であるとき、

$$E(xy) = E(x)E(y) \quad \dots\dots\dots (12)$$

が成立する。ここで、「互いに独立であるとき」の意味であるが、 x と y が影響を及ぼしあうことはなく、 x, y の値がそれぞれ別個にある確率に従って決定する、ということを表している。二つのサイコロを同時に振ったとき、それぞれ出た目は独立であると言えるだろう。あるサイコロの目はもう一つのサイコロの目が何であるかということとは関係なく決定することは自明である。このときもサイコロを例に取り計算してみるとこの式の意味がよくわかるだろう。

4・2 母分散について

4・1 では期待値の基礎について考えたが、この応用をここでは考えよう。ばらつきは分散で表されると前章で解説したが、これを期待値で表すことを考える。

分散とは測定値と平均値の二乗平均である。つまりこれを期待値で表すと、

$$\text{var}(x) = E\{(x - \mu)^2\} \quad \dots\dots\dots (13)$$

となる。 $\text{var}(x)$ は x の分散を表している。また式(13)では、測定値 x と母平均 μ との偏差の二乗の期待値を求めている。つまり $\text{var}(x)$ は母分散を表している。また、先ほど言ったように母集団の性質を表す変数はギリシャ文字が用いられる。母分散は通常 σ^2 で表される。

この分散もいくつか便利な性質を持っている。その性質をいくつか挙げる。

c が定数のとき、

$$\text{var}(x + c) = \text{var}(x) \quad \dots\dots\dots (14)$$

が成立する。これは、例えばサイコロの目を1~6から3~8に変えたところで、平均値は変わるがばらつきの

大きさは変わらないということを表している。

c が定数のとき、

$$\text{var}(cx) = c^2 \text{var}(x) \quad \dots\dots\dots (15)$$

が成立する。これは、

$$\begin{aligned} \text{var}(cx) &= E[\{c(x - \mu)\}^2] = E\{c^2(x - \mu)^2\} \\ &= c^2 E\{(x - \mu)^2\} = c^2 \text{var}(x) \end{aligned}$$

だからである。

x, y が独立のとき、

$$\text{var}(x \pm y) = \text{var}(x) + \text{var}(y) \quad \dots\dots\dots (16)$$

が成立する。これは、

$$\begin{aligned} \text{var}(x \pm y) &= E[\{(x - \mu_x)^2 \pm (y - \mu_y)^2\}] \\ &= E\{(x - \mu_x)^2 + (y - \mu_y)^2 \pm 2(x - \mu_x)(y - \mu_y)\} \\ &= E\{(x - \mu_x)^2\} + E\{(y - \mu_y)^2\} \\ &\quad \pm 2E\{(x - \mu_x)(y - \mu_y)\} \end{aligned}$$

となる。ここで、 $E\{(x - \mu_x)(y - \mu_y)\}$ という項があるが、この項は x, y が互いに独立であれば0となる性質を持つ。そうすると、

$$\begin{aligned} \text{var}(x \pm y) &= E\{(x - \mu_x)^2\} + E\{(y - \mu_y)^2\} \\ &= \text{var}(x) + \text{var}(y) \end{aligned}$$

となる。式(16) は大変重要である。つまり、ばらつきを合成するときには分散をそのまま足し算すればよい、もしくは、標準偏差を合成するときには二乗和の平方根を用いればよい、ということを表している。誤差、不確かさの合成はこの性質を用いて行っている。

5 平均値の分散

通常、測定結果は標本平均を最終的な測定結果として報告する。そうすると重要になるのが平均値の分散である。標本平均はデータを取得するたびに異なる値が算出される(たまたま同じ値になるときもあるが)。これもサイコロを考えれば自明であろう。サイコロを5回振ったときの平均値は毎回同じ平均値になるわけではなく、5回振るたびに異なる平均値が求められる。つまり、標本平均とは変動する値なのである。この標本平均のばらつきを表したものが平均値の分散である。

標本平均は、

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1}{n}(x_1 + x_2 + x_3 + \dots + x_n) \quad \dots\dots (17)$$

によって算出される。よって平均値の分散は、

$$\begin{aligned} \text{var}(\bar{x}) &= \text{var}\left\{\frac{1}{n}(x_1 + x_2 + x_3 + \dots + x_n)\right\} \\ &\dots\dots\dots (18) \end{aligned}$$

で表される。ここで n は定数であるので、式(15) より、

$$\text{var}(\bar{x}) = \frac{1}{n^2} \text{var}(x_1 + x_2 + x_3 + \cdots x_n) \quad \cdots \cdots (19)$$

となる。また、各 x_i は繰り返し測定より得られることがほとんどであろう。このようなとき、各 x_i は他の x_i の値に影響されることはなく値が決定すると考えても、そうおかしい前提をおいているわけではない。つまり各 x_i は互いに独立であるとする、式(16) から、

$$\text{var}(\bar{x}) = \frac{1}{n^2} \{\text{var}(x_1) + \text{var}(x_2) + \cdots + \text{var}(x_n)\} \quad \cdots \cdots (20)$$

となる。ここで各 x_i の分散を考える。各 x_i は、それぞれ同じ測定から得られている。すなわち同じ母集団からサンプリングされた標本である。そうであれば、 x_i の分散は母分散である $\sigma^2(x)$ と考えられる。式で示すと、

$$\text{var}(x_1) = \text{var}(x_2) = \cdots = \text{var}(x_n) = \sigma^2(x) \quad \cdots (21)$$

となる。これを式(20) に代入すると、

$$\begin{aligned} \text{var}(\bar{x}) &= \frac{1}{n^2} \{\text{var}(x_1) + \text{var}(x_2) + \cdots + \text{var}(x_n)\} \\ &= \frac{1}{n^2} \{\sigma^2(x) + \sigma^2(x) + \sigma^2(x) + \cdots + \sigma^2(x)\} \\ &= \frac{1}{n^2} \{n\sigma^2(x)\} \\ \sigma^2(\bar{x}) &= \frac{\sigma^2(x)}{n} \quad \cdots \cdots (22) \end{aligned}$$

となる。つまり、母集団から n 個サンプリングして算出した標本平均の分散は母分散の $1/n$ になるということを表している。

この平均値の分散の性質を考えると、平均値を算出するためのデータが増えれば増えるほど標本のばらつきは小さくなる。逆に、測定を1回しか行わないのであれば、その1回の測定で得られたデータがそのまま標本平均となる。ある一つのデータの分散はもちろん σ^2 となるが、これは式(22) の n に1を代入したものと等しい。

式(22) を見てわかるように、これは母分散に関する式である。母分散は無限回の測定を行わなければ算出できない値であるので、我々には知ることができない。よって実際に計算を行うときには母分散 $\sigma^2(x)$ の代わりに母分散の推定値 $\hat{\sigma}^2(x)$ を用いる必要がある。ここで、“ $\hat{}$ ” のついたギリシャ文字は母数の推定値を表す。よって式(22) は、

$$\hat{\sigma}^2(\bar{x}) = \frac{\hat{\sigma}^2(x)}{n} \quad \cdots \cdots (23)$$

となる。そして母分散の推定値として通常用いられるのは標本分散 $s^2(x)$ である。式(23) は平均値の推定・検定、不確かさ評価など非常に多くの場所で用いられる。

6 自由度と不偏推定量

第2章で標本分散について見てきたが、標本分散では偏差の二乗平均を算出するときにデータの個数ではなく(データの個数 - 1) で表される自由度で割った。これはなぜだろうか？

ここでは、新しい分散 $\tilde{s}^2(x)$ を考えよう。これは、

$$\tilde{s}^2(x) = \frac{\sum_i (x_i - \bar{x})^2}{n} \quad \cdots \cdots (24)$$

という、自由度ではなくデータの個数 n で割った分散である。この $\tilde{s}^2(x)$ はいったいどのような値を推定しているのかを第4章で解説した期待値を用いて考える。

まず式(24) を変形する。

$$\begin{aligned} \tilde{s}^2(x) &= \frac{\sum_i (x_i - \bar{x})^2}{n} = \frac{\sum_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n} \\ &= \frac{\sum_i x_i^2}{n} - 2\bar{x} \frac{\sum_i x_i}{n} + \frac{n\bar{x}^2}{n} \\ &= \frac{\sum_i x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 = \frac{\sum_i x_i^2}{n} - \bar{x}^2 \quad \cdots \cdots (25) \end{aligned}$$

式(25) の期待値を求めてみよう。

$$\begin{aligned} E[\tilde{s}^2(x)] &= E\left(\frac{\sum_i x_i^2}{n} - \bar{x}^2\right) \\ &= \frac{1}{n} \sum_i \{E(x_i^2)\} - E(\bar{x}^2) \quad \cdots \cdots (26) \end{aligned}$$

ここで、 $E(x_i^2)$ とは何を表しているのかを考える。式(13) より、

$$\text{var}(x) = E\{(x - \mu)^2\} = E(x^2 - 2x\mu + \mu^2) \quad \cdots \cdots (27)$$

ここで μ は母平均であるので定数である。よって式(27) は、

$$\begin{aligned} E(x^2 - 2x\mu + \mu^2) &= E(x^2) - 2\mu E(x) + \mu^2 \\ &= E(x^2) - 2\mu \times \mu + \mu^2 = E(x^2) - \mu^2 \quad \cdots (28) \end{aligned}$$

となる。 $\text{var}(x) = \sigma^2$ であることに留意すると、

$$E(x^2) = \sigma^2 + \mu^2 \quad \cdots \cdots (29)$$

となる。また $E(\bar{x}^2)$ も同様に考えると、

$$\begin{aligned} \text{var}(\bar{x}) &= E\{(\bar{x} - \mu)^2\} = E(\bar{x}^2 - 2\bar{x}\mu + \mu^2) \\ &= E(\bar{x}^2) - 2\mu E(\bar{x}) + \mu^2 \\ &= E(\bar{x}^2) - 2\mu \times \mu + \mu^2 \\ &= E(\bar{x}^2) - \mu^2 \end{aligned}$$

となり, $\text{var}(\bar{x}) = \sigma^2/n$ であることに留意すると,

$$E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2 \dots\dots\dots (30)$$

となる。式(29) と式(30) を式(26) に代入すると,

$$\begin{aligned} E[\bar{s}^2(x)] &= \frac{1}{n} \sum (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2 \right) \\ &= \frac{1}{n} (n\sigma^2 + n\mu^2) - \left(\frac{\sigma^2}{n} + \mu^2 \right) \\ &= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \sigma^2 - \frac{\sigma^2}{n} \end{aligned}$$

よって,

$$E[\bar{s}^2(x)] = \frac{n-1}{n} \sigma^2 \dots\dots\dots (31)$$

となる。つまり, 自由度ではなく測定値の個数 n で偏差の二乗和を割って算出した分散は母分散を推定しているのではないことがわかる。また式(31) から,

$$E \left\{ \frac{\sum_i (x_i - \bar{x})^2}{n-1} \right\} = \sigma^2 \dots\dots\dots (32)$$

であることは自明であろう。つまり標本分散 $s^2(x)$ は母平均の推定値として全く偏りをもたないということである。このような母数の推定量のことを不偏推定量と呼ぶ。つまり, データの個数ではなく自由度で偏差の二乗和を割ったのは不偏推定量にするためなのである。よって, この $s^2(x)$ のことを不偏分散とも呼ぶ。

また, 標本平均も,

$$E(\bar{x}) = E \left(\frac{\sum_i x_i}{n} \right) = \frac{1}{n} \sum_i E(x_i) = \frac{1}{n} \sum_i \mu$$

$$= \frac{1}{n} \times n\mu$$

$$E(\bar{x}) = \mu \dots\dots\dots (33)$$

ということから不偏推定量である。

統計は標本平均, 標本分散を基本とした体系で構成されているが, これは標本平均, 標本分散が不偏推定量であることが一番大きな理由である。また, 標本標準偏差は不偏分散の平方根であるが, 母標準偏差の不偏推定量ではない。

7 最後 に

第1回は統計の基礎ということで話を進めたが, 特に「2. 測定と統計」を重視して欲しい。測定を行う前にはどのような統計的手法を適用するのかということが決定しているのであれば, 実験データの取得の際に明確な目的意識を持てる。これは質のよいデータを取得するとき大変重要なことである。統計の数学的な話も重要ではあるが, 統計的手法を当てはめるためのデータの質が低ければいくらすばらしい統計的手法を用いることができてても妥当な結果が得られることはない。