

連載 (講義)

数値データの取り扱い ～四捨五入から検出限界まで～ 第二回 平均値, 区間推定と棄却検定

福島 整

独立行政法人物質・材料研究機構 共用基盤部門分析支援ステーション

〒305-0047 つくば市千現 1-2-1

fukushima.sei@nims.go.jp

(2010年10月2日受理)

分析業務にとって、数字の取り扱いは命である。有効数字やデータのばらつきは、ほとんどの現場担当者が無意識のうちに正当な方法で処理しているはずであるが、あらためて統計学の用語を目の前にすると自信を無くす人が多い。本講は、そのような人たちに自信を持たせ、あるいは体系的な知識を学習するための出発点と成ることを意図したものである。今回は、平均値, 区間推定の概念及び棄却検定の基礎的な部分について説明する。

The handling of numeric data

- from “round off” to “detection limit”-

II. The basis of average, interval estimation, and rejection test

Sei FUKUSHIMA

Materials Analysis Station, Dept. Mat. Infrastructure, NIMS, 1-2-2 Sengen, Tukuba 305-0047, Japan

fukushima.sei@nims.go.jp

(Received: October 2, 2010)

One of the most important items for the analyst is the handling of the numerical data. The almost analysts handle and treat the numeric data, significant figure or dispersion for example, obtained from the measurement in the correct way unconsciously. However, there are many analysts who lose the self-confidence, when they wish to use the description of the statistical terminology. This lecture is intended giving a starting point for giving the self-confidence and learning a systematic knowledge to such people. In present part of this lecture, the basis of average, interval estimation, and rejection test will be explained.

4. 平均値

4. 1 平均値とは

さて今回は、分析業務でもっともふつうに用いられる平均値について考えてみよう。データの集まりを $\{x_1, x_2, \dots, x_n\}$ とした場合、通常平均値 \bar{x} は

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (22)$$

とあらわされる。しかし、平均の取り方はこれだけで

はない。(22)は、別に「単純平均」とか「算術平均」「相加平均」と呼ばれる平均値であるが、あらゆる場合に適切な(「正しい」ではない事に注意)母集団の平均値の推定値を与えるわけではないのである。ましてや、常に真値に近い値を与えるわけでもない。

単純平均以外で比較的なじみの深い平均の取り方として、「二乗平均」がある。二乗平均の計算は、すでに出てきた分散の計算とほぼ同じ式である。すなわち、

データの集まり $\{x_1, x_2, \dots, x_n\}$ に対して

$$x_{rms} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} \quad (23)$$

で計算される. 二乗平均は別名「二乗平均平方根」とか「平均二乗偏差」「RMS (Root Mean Square)」とも呼ばれ, 単純平均とは少々違う使われ方をする. すなわち, 「偏差」という用語が示すとおり, 「真の値」からどの程度ばらついているかという指標を与えるのである. x_{rms} は \bar{x} とその分散 σ^2 を用いると

$$x_{rms}^2 = \bar{x}^2 + \sigma^2 \quad (25)$$

とあらわされる. すなわち, 常に単純平均より大きな値を示すとともに, 単純平均が 0 のときにはそのデータの分散を与える事がわかる. このことから, 真の値からのばらつきを表しているということが, なんとなく想像がつくであろう.

あるいは「相乗平均」も, 不等式の証明問題の一つとして数学の教科書で見たことを覚えておられるかもしれない. 相乗平均は英語で Geometric mean と呼ばれ, 日本語でも「幾何平均」という別名がある. 相乗平均 x_{geo} は,

$$x_{geo} = \sqrt[n]{x_1 x_2 \dots x_n} \quad (26)$$

で与えられる. x_i が, i について順序性があるときなどにつかわれる. 例えば, ある日の体積膨張が 20% で次の日が 80% だったとすると, 2 日間の膨張率の平均は $\sqrt{1.2 \cdot 1.8} \approx 1.47$ となり, 約 47% と計算できる. (例えば $\frac{1.2 \cdot 1.8 - 1}{2} = 0.58$ だから 58% と答えると間違いである. すなわち, $1.58^2 = 2.4964 \neq 1.2 \cdot 1.8$ だからである.)

また, 「調和平均」(Harmonic mean) と呼ばれる平均値も使用される. 調和平均は, 以下の式

$$x_{har} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} \quad (26)$$

で計算される. もっともよく使われる例は, 電気回路において並列に接合された抵抗器全体で示す抵抗値の算出であろう. 同じ基準で表された割合同士の平均を求めるときに用いられるのが, 調和平均である.

平均速度の問題も, 調和平均の一つの良い例となる. 例えば, 行きが毎分 60m, 帰りが毎分 90m のスピードで往復したときの平均スピードは,

$$\frac{2}{\frac{1}{60} + \frac{1}{90}} = 72$$

となり, 毎分 72m となる. (例えば, $\frac{60+90}{2} = 75$ で毎分 75m と答えると間違いとなる. すなわち, 仮に 180m の距離を往復したとすると, 行きに 3 分, 帰りに 2 分,

合計 5 分かかったことになるから, $\frac{180 \times 2}{5} = 72$ となる.) 数学の世界では, これらを統合した一般化平均

$$x_{genral} = \sqrt[m]{\frac{x_1^m + x_2^m + \dots + x_n^m}{n}} \quad (27)$$

が定義されている. それぞれ $m=1, m=-1$ が単純平均及び調和平均となり, $m \rightarrow 0$ の極限で相乗平均を表せることがわかる.

このように, 平均値と一口に言っても様々な種類があり, 目的によって使い分けることになる.

4. 2 単純平均の性質 (不偏推定)

出だしから少々脱線気味であるが, 実際には(22)の単純平均がもっとも実用的に用いられているのは言うまでもない. この単純平均は, 統計学上大変都合の良い性質を持つ. すなわち, 「母集団が計算可能な平均値と分散を持つとき, 母平均に対してもっとも良い不偏推定量」(正確には「線形不偏推定量」)なのである. 母集団は, 正規分布でなくてよい. また, 通常で計算された分散も, 母集団に対する不偏推定量である.

この「不偏推定量」とは, 推定しようとする値に最も近い数値である(偏りが少ない, あるいは無い)ことを意味する. すなわち単純平均は, その値が得られる元となるデータの範囲で, 母平均に最も近い一致した値を与えるのである. また分散は, 母分散の $(n-1)/n$ で与えられることを前回の最後に示したが, これが不偏推定量に相当する.

以下に, 多少数学を駆使して単純平均が最も良い「線形不偏推定量」であることを示す. 数式の展開を理解する必要はないが, 単純平均といえども厳密な理論の展開の裏付けを有していることを実感して頂ければと考える.

さて, 前回の講義の最後に期待値を与える関数を用いた説明を行った. ここでの議論でも期待値を用いるので, その関数 $E(x)$ についてきちんと定義を示そう. x はデータ $\{x_1, x_2, \dots, x_n\}$ をとるものとし, それぞれの値をとる確率が $\{p_1, p_2, \dots, p_n\}$ であるとしよう. ここで, $p_1 + p_2 + \dots + p_n = 1$ である.

すると, x の期待値は

$$E(x) = \sum_{i=1}^n x_i p_i \quad (28)$$

で定義される. この関数 $E(x)$ は, 次のような性質を持つ. すなわち a と b を定数とすれば

$$E(ax \pm by) = aE(x) \pm bE(y) \quad (29)$$

次に, $\{x_1, x_2, \dots, x_n\}$ それぞれの値に対応した別の

連の値 $\{w_1, w_2, \dots, w_n\}$ を仮定して, 次式の様な値 \tilde{x} を作る.

$$\tilde{x} = \sum_{i=1}^n w_i x_i \quad (30)$$

(30)の計算は, x_i に重み w_i をかけて足し合わせることを意味しており, この様な計算を x_i の「線形結合」と呼ぶ. \tilde{x} を一種の推定値と見なすことにし, これがもし「不偏推定量」であるとすれば, $\{x_1, x_2, \dots, x_n\}$ の母集団の平均値を \bar{X} とすると, \tilde{x} の期待値 $E(\tilde{x})$ が

$$E(\tilde{x}) = \bar{X} \quad (31)$$

となっていることが, 「不偏推定量」であることの定義である. また, 「不偏推定量」であるために, $w_1 + w_2 + \dots + w_n = 1$ でなければならない. 要するに, w_i は(28)の p_i と対応していて, それぞれの値をとる確率と考える事と実質的に同じである. このあたりの詳しい理由を知りたい方は, 統計の教科書の「不偏推定」の項を調べて頂きたい.

さて問題は, \tilde{x} が \bar{X} に対して最も良い(最も差の少ない)推定量となるためには, 重み w_i がどのような値をとるかを調べる事である. それには, \tilde{x} がどの程度母集団の平均値 \bar{X} からばらつくかを求める式を作り, その性質を見てみればよい. それには, \tilde{x} の期待値が母集団の平均値 \bar{X} と等しいとした定義と w_i の総和が1であることを利用して, (30)で定義された推定値と母集団の平均値の差をとり

$$\tilde{x} - E(\tilde{x}) = \sum_{i=1}^n w_i x_i - \bar{X} = \sum_{i=1}^n w_i (x_i - \bar{X}) \quad (32)$$

ばらつきは「平均値との差の二乗の期待値」であることが統計学の教科書に記されているので, その定義を利用し, \bar{X} が定数であるので $E(\bar{X}) = \bar{X}$ であることに注意すれば

$$\begin{aligned} E[(\tilde{x} - E(\tilde{x}))^2] &= E\left[\left(\sum_{i=1}^n w_i (x_i - \bar{X})\right)^2\right] \\ &= E\left(\sum_{i=1}^n \sum_{j=1}^n w_i w_j (x_i - \bar{X})(x_j - \bar{X})\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j E((x_i - \bar{X})(x_j - \bar{X})) \end{aligned} \quad (33)$$

ここで, 各 x_i は互いに無関係であるとするならば, $i \neq j$ のときは

$$E((x_i - \bar{X})(x_j - \bar{X})) = 0 \quad (34)$$

であるというのが統計学の教えるところであり(互いに無関係ならば共分散は0に等しい), また \bar{X} は母集団の平均値であるから定数あることから, $E((x_i - \bar{X})^2)$ も定数として扱えるから, それを c とす

れば, (33)は

$$E[(\tilde{x} - E(\tilde{x}))^2] = c \sum_{i=1}^n w_i^2 \quad (35)$$

という極めて簡単な式に書き直せる. したがってこの問題は, 結局 w_i の総和が1である($\sum_{i=1}^n w_i = 1$)という拘束の下で(35)の値が最小になる w_i を見つければよいことになる. これには, 解析学のツールの一つであるラグランジュの未定係数法を用いる. すなわち(35)と拘束条件 $\sum_{i=1}^n w_i = 1$ から関数 F を以下のようにとり

$$F = \sum_{i=1}^n w_i^2 + \lambda \left(1 - \sum_{i=1}^n w_i\right) \quad (36)$$

F を最小にする w_i を求めればよい. したがって, F を w_i で微分して0とおけば

$$\begin{aligned} \frac{\partial F}{\partial w_i} &= 2w_i - \lambda = 0 \\ \therefore w_i &= \frac{\lambda}{2} \end{aligned} \quad (37)$$

拘束条件 $\sum_{i=1}^n w_i = 1$ を用いると

$$\begin{aligned} \sum_{i=1}^n w_i &= \frac{1}{2} \sum_{i=1}^n \lambda = \frac{n\lambda}{2} = 1 \\ \therefore \lambda &= \frac{2}{n} \\ \therefore w_i &= \frac{1}{n} \end{aligned} \quad (38)$$

となる. つまり, n 個のおもみが総て同じ $1/n$ をとるときに, 線形推定量 \tilde{x} は母集団の平均値 \bar{X} に最も近くなる. すなわち不偏推定量を与えるのである.

繰り返すが, 単純平均を用いる時に, この様なめんどくさい理論を理解する必要はない. むしろ, ここで展開したような厳密な理論に裏打ちされているので, 最も性質の良い推定値として安心して利用できるのだと言うことを実感して頂ければ, 十分である.

なお, 不偏推定の理論自体は結構ややこしいことも, 以上の展開で御理解頂けたと思う. 興味のある方は, 是非, 教科書を参照して頂きたい.

4. 3 データから値を推定するとは(区間推定)

さて, 通常我々が用いている平均値(単純平均)の正体ははっきりしたが, 一つだけ注意しておかねばならないことがある. それは, 「平均値を真の値と誤解してはならない」ということである.

平均値は, 先の説明の通り「母平均の不偏推定量」である. しかし,

母平均は真の値ではない.

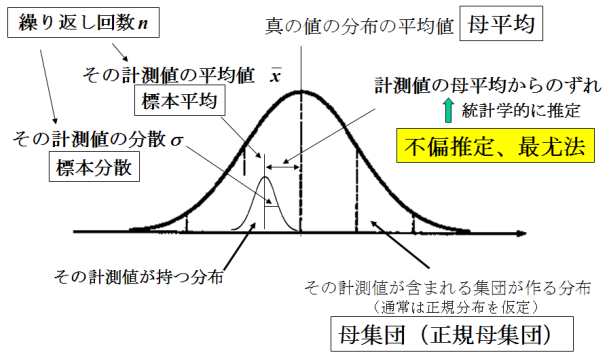


図6 標本平均と標本分散の母集団との関係

統計学では、「母平均を、真の値の代わりになると仮定」して、議論を進めているだけである。これを再度ここで確認しておこう。

前回の講義で、計測値は、母集団を仮定しそこから無作為に抜き取った値と見なすことが出来ると説明した。それにより、標本平均を単純平均値として求めると、母集団の平均値 (母平均) に最も近い値を与えてくれるというのが、図6にまとめてもある前節の説明の内容であった。

しかし実際には、元のデータがどのような分布に属しているかなど、分かりっこないのである。したがって、とりあえず正規分布を仮定し、測定はそこからデータを取り出す (標本を抽出する) 事と仮定し、平均値と分散を計算しているのだ (図7)。その値が与える情報とは、図7の下の方に示してある灰色の区間なのである。例えば大体 68%の確率で母平均の値が含まれている「値の存在区間」は $(x - \sigma, x + \sigma)$ で与えられる。98%以上を期待するのであれば、「値の存在区間」は $(x - 3\sigma, x + 3\sigma)$ をとればよい。

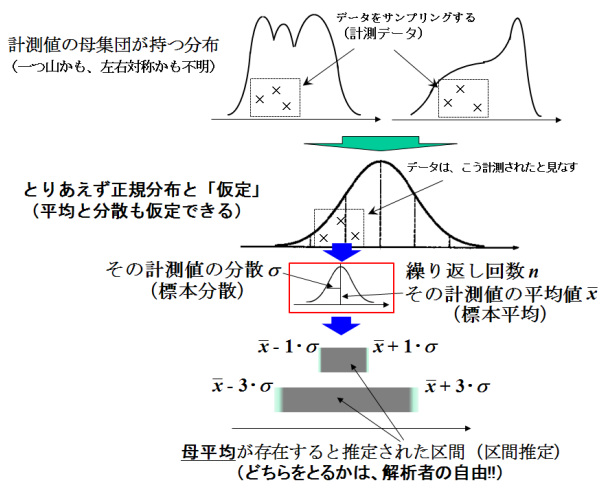


図7 区間推定のおおざっぱな概念

ここで、ちょっと注意して頂きたいのは、値の区間を記述するための書き方である。「値 a と b に挟まれた区間」を記述するのに、 (a, b) と $[a, b]$ という2つの書き方がある。 (a, b) は、この区間の両側の値 a と b は、この区間に含まれない。一方、 $[a, b]$ と書くと、両側の値 a と b もこの区間に含まれる。すなわち、例えば $(x - \sigma, x + \sigma)$ の場合であれば、 $x - \sigma$ と $x + \sigma$ は、推定区間に含まれないのである。

当然、区間を広くとれば、その中に「当たり」が含まれる確率は高くなる。この確率は、解析者が自分で判断して決めねばならない。そして、この区間に含まれる総ての値は、「どれも等しく母平均の候補」なのだ。平均値を特別扱いする根拠は、何もないのである。

このように、値の存在範囲を推定することを区間推定と呼ぶ。これに対して、平均値や分散のように特定の値を求めることを点推定と呼ぶ。正確に定義し直すと、「点推定」とは「標本集団の値を用いて、母集団の分布を表現するパラメータを数値として推定すること」である。これに対して、「区間推定」とは「点推定で推定したパラメータの分散や信頼区間を示すこと」である。図7の下に示したものは、標本平均の信頼区間であり、その意味はすでに説明したとおりである。

では、ここまでご説明したようなややこしい概念をなんにも考えずとも、平均値をあたかも真の値同様に使用した結論を用いてもほとんど不都合が起こっていないのは何故か。筆者が思うに、とどのつまりは、自然は親切であり、単純なのだ。計測者がよほど変なことをしない限り (整備の悪い装置を使用するとか)、データはほぼ正規分布的なのだ。...というか、多分そうなのだろう。

5. 検定

5.1 検定とは

さて、平均値を求める為に複数のデータを使用する事になるわけであるが、手元のデータを何も考えずにそのまま使って良いのかどうか迷ったことはないだろうか。なるべく繰り返しの実験を頑張って、何個かの結果の数字が得られたとする。その中に、一つだけ飛び抜けて離れた値が存在していたりすると、そのデータを捨てるのは忍びないが、それを入れると平均値に対する標準偏差もあまり良くないし...と悩むことはよくある。

確かに、なるべく「たちの良い」平均値を得ることが、後の処理の信頼性も高いものとはできることは、容易に想像がつこう。しかし、数回の繰り返し測定を行

うと、どうしても1個ぐらいかけ離れて見える結果が得られるものである。「常に正しい測定がなされている」という大前提が満足されているのであれば(飛び離れたデータの原因が合理的に推定できない場合), このかけ離れた数字を統計学的な仮定の下に検討し, 以後の処理に使わない(データの棄却)かどうかを決めることができる。

また, 例えば同じ測定を同じ装置で繰り返す実験を, 実施期間を改めて2回行ったとする。その場合, 2セットのデータと平均値及び分散が得られるわけであるが, これらを併せて使用して良いかどうか悩んだこと無いだろうか。この判断にも, 検定を用いることが出来る。

統計の教科書をひもといてみると, 検定とは「区間推定値から, 母集団が特定の分布に従っているかどうかを検証すること」とある。すなわち, 前章で説明した区間推定が, ここでは重要な役割を果たすのだ。

実際にどうするかと言うと, 標本(測定データと思えばよい)が特定の分布に従う母集団から抽出されたとする仮説を立て, この仮説の真偽を調べるのである。これが, 「検定」である。つまり, 前節までの議論では, 測定値がある分布を持つ母集団から抜き出された物だという考えで, 平均や分散を取り扱った。この母分散を利用し, あらたに現れたデータがこの分散に従っているかどうか見極めようとするのが検定だということになる。

検定の結果は, yes か no である。この判断をするために, まず仮説を立てる。例えば, 「新たに得られたデータが, 確率0%で母集団に属するものである」というようなものである。この最初に立てられる仮説を, 統計では「帰無仮説(きむかせつ)」という。次に, こうした帰無仮説から予想される値(理論的な分布から求めたりする)と, 実際に測定データから計算された値が一致する確率(よく「p値」と呼ばれる)を求め, 帰無仮説から求めたり, 検定のためにデータから計算されたりする量は, 「統計量」と呼ばれる。

このp値が, 帰無仮説にしたがって決めた基準(「有意水準」と呼ぶ。αで表されることが多く, 5%または1%が使用されることが多い)よりも小さい(つまり「起こりそうもない」)場合には「有意差がある」として, 帰無仮説は棄却される。そうでなければ, 「有意差が無い」として, 帰無仮説が採用される。

検定を行うためには, まず, 帰無仮説から予想される量の存在範囲を決めねばならない。これには, 2つのやり方がある。

一つは, 区間推定に用いられたパラメータ(平均値

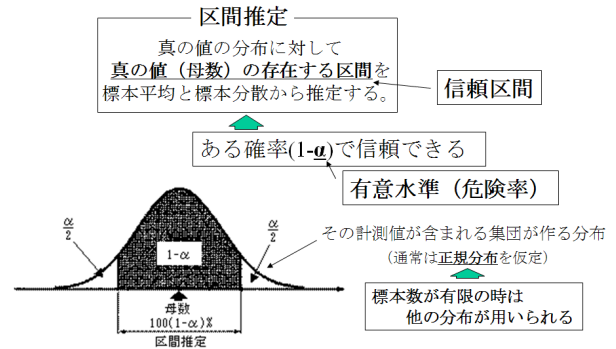


図8 区間推定と信頼区間の関係

や分散)を用いて理論的に決めるやり方である。これを, パラメトリック検定という。特に, pによって決めた母平均が存在する区間を「信頼区間」と呼ぶ。また, 実際にはデータの個数が有限個であることから, この影響を補正するための新たな理論関数により信頼区間を推定することもなされる。

また, 信頼区間が前もって数表等で与えられており, 解析者が式に従ってデータから統計量を計算し有意水準に基づいて判断する検定もある。この代表的なものの一つが棄却検定であり, 突発データのように一つだけ飛び離れているように見えるデータをどうするか判断に有効である。管理図による品質管理も, 広義の棄却検定と言える。

Grubbsの方法

1つだけ飛び抜けた値がある場合

- まず, データ $x_1 \leq x_2 \leq \dots \leq x_n$ に対して $\bar{x} = \frac{\sum x_i}{n}$ かつ $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ とした場合, 最大あるいは最小の値に対して T_1 あるいは T_n を求める。

$$T_1 = \frac{\bar{x} - x_1}{\sigma} \quad T_n = \frac{x_n - \bar{x}}{\sigma}$$
- 有意水準 α (間違えて棄ててしまう確率) を決める。

通常 α=0.05 あるいは 0.01
(間違えて棄ててしまう確率が5%あるいは1%)
- 有意水準 α とデータ数 n から対応する値 T を表から探し出す

$$T_1 \text{ あるいは } T_n \text{ が } T \text{ の値以上である}$$

$$T_1, T_n \geq T$$
- 対応する x_1 あるいは x_n が, 間違える確率 α のもとで異常値と判断できる。

・ σ が過去のデータの蓄積より得られた場合にも, (σ が母分散と見なされるとき) 対応する棄却検定表が存在。

図9 Grubbsの棄却検定の手順

5. 2 棄却検定

例えば、一連の n 個の測定データについて、例えば $x_1 \leq x_2 \leq \dots \leq x_n$ のように並べ替えられたとすると、多くの場合最大のデータか最小のデータが異常値として問題となる。最大か最小のデータ 1 個について、棄却できるかどうかの判別法として、ISO 5725/JIS Z 8402 では Grubbs の方法[1]をまず推奨している。この方法では、まず総てのデータ n 個に対して、平均 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ と分散 $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$ を計算する。これを用い、 x_1 に対して

$$T_1 = \frac{\bar{x} - x_1}{\sigma} \tag{39}$$

あるいは x_n に対して

$$T_n = \frac{x_n - \bar{x}}{\sigma} \tag{40}$$

を求めておく。あわせて、間違えて捨ててしまう判断をする可能性（有意水準）を確率の数字で与える。この値は、 $\alpha = 0.05$ あるいは 0.01 （間違える確率が 5% あるいは 1%）が推奨されている。Grubbs は、 α 及びデータ数 n に対して、捨てていいかどうかの基準を与える数表を示している。(39)もしくは(40) で与えられる値が数表の値以上であれば、その間違える確率（危険率）を覚悟の上でデータを捨てる判断をとることになる。

表 1 Grubbs の棄却検定表

| データ数 n | 上側確率(α) | | | |
|-------------|------------------|-------|-------|-------|
| | 0.05 | 0.025 | 0.01 | 0.005 |
| 3 | 1.153 | 1.155 | 1.155 | 1.155 |
| 4 | 1.463 | 1.418 | 1.492 | 1.496 |
| 5 | 1.672 | 1.715 | 1.749 | 1.764 |
| 6 | 1.822 | 1.887 | 1.944 | 1.973 |
| 7 | 1.938 | 2.020 | 2.097 | 2.139 |
| 8 | 2.032 | 2.126 | 2.221 | 2.274 |
| 9 | 2.110 | 2.215 | 2.323 | 2.387 |
| 10 | 2.176 | 2.290 | 2.410 | 2.482 |
| 11 | 2.234 | 2.355 | 2.485 | 2.564 |
| 12 | 2.285 | 2.412 | 2.550 | 2.636 |
| 13 | 2.331 | 2.462 | 2.607 | 2.699 |
| 14 | 2.371 | 2.507 | 2.659 | 2.755 |
| 15 | 2.409 | 2.549 | 2.705 | 2.806 |
| 16 | 2.443 | 2.585 | 2.747 | 2.852 |
| 17 | 2.475 | 2.620 | 2.785 | 2.894 |
| 18 | 2.504 | 2.651 | 2.821 | 2.932 |
| 19 | 2.532 | 2.681 | 2.854 | 2.968 |
| 20 | 2.557 | 2.709 | 2.884 | 3.001 |
| 21 | 2.580 | 2.733 | 2.912 | 3.031 |
| 22 | 2.603 | 2.758 | 2.939 | 3.060 |
| 23 | 2.624 | 2.781 | 2.963 | 3.087 |
| 24 | 2.644 | 2.802 | 2.987 | 3.112 |
| 25 | 2.663 | 2.822 | 3.009 | 3.135 |
| 26 | 2.681 | 2.841 | 3.029 | 3.157 |
| 27 | 2.698 | 2.859 | 3.049 | 3.178 |
| 28 | 2.714 | 2.876 | 3.068 | 3.199 |

| データ数 n | 有意水準 | | | 計算式 | |
|-------------|-------|-------|-------|--|--|
| | 0.10 | 0.05 | 0.01 | | |
| 3 | 0.886 | 0.941 | 0.988 | 最小値が疑わしい: $r_{10} = (x_2 - x_1)/(x_n - x_1)$ 最大値が疑わしい: $r_{10} = (x_n - x_{n-1})/(x_n - x_1)$ | |
| 4 | 0.679 | 0.765 | 0.889 | | |
| 5 | 0.557 | 0.642 | 0.780 | | |
| 6 | 0.482 | 0.560 | 0.698 | | |
| 7 | 0.434 | 0.507 | 0.637 | | |
| 8 | 0.479 | 0.554 | 0.683 | | 最小値が疑わしい: $r_{10} = (x_2 - x_1)/(x_n - x_1)$ 最大値が疑わしい: $r_{10} = (x_n - x_{n-1})/(x_n - x_2)$ |
| 9 | 0.441 | 0.512 | 0.635 | | |
| 10 | 0.409 | 0.477 | 0.597 | | |
| 11 | 0.517 | 0.576 | 0.679 | 最小値が疑わしい: $r_{21} = (x_3 - x_1)/(x_{n-1} - x_1)$ 最大値が疑わしい: $r_{10} = (x_n - x_{n-2})/(x_n - x_2)$ | |
| 12 | 0.490 | 0.546 | 0.642 | | |
| 13 | 0.467 | 0.521 | 0.615 | | |
| 14 | 0.492 | 0.546 | 0.641 | 最小値が疑わしい: $r_{22} = (x_3 - x_1)/(x_{n-2} - x_1)$ 最大値が疑わしい: $r_{10} = (x_n - x_{n-2})/(x_n - x_3)$ | |
| 15 | 0.472 | 0.525 | 0.616 | | |
| 16 | 0.424 | 0.507 | 0.595 | | |
| 17 | 0.438 | 0.490 | 0.577 | | |
| 18 | 0.424 | 0.475 | 0.561 | | |
| 19 | 0.412 | 0.462 | 0.547 | | |
| 20 | 0.401 | 0.450 | 0.535 | | |
| 21 | 0.391 | 0.440 | 0.524 | | |
| 22 | 0.382 | 0.430 | 0.514 | | |
| 23 | 0.374 | 0.421 | 0.505 | | |
| 24 | 0.368 | 0.413 | 0.497 | | |
| 25 | 0.360 | 0.406 | 0.489 | | |

・有意水準とデータ数 n から、表の対応する式を用いて得られた統計量を左の表の値と比較し、大きければ異常値と判断する。

図 10 Dixon の方法

もし、 x_1 と x_n のどちらかが異常値かあらかじめ予想できない場合は、 α の値を倍にして判断する（有意水準 0.05 であれば、 $\alpha = 0.1$ とする）。以上のプロセスを図 9 に、棄却検定のための数表を表 1 に示す。多くの統計学の教科書に、この表は載せられている。

また、もっと簡便な方法として Dixon の方法[2]も利用できる。計算方法と数表の例を、図 10 に示しておく。この方法は、ISO 5725/JIS Z 8402 に提示されているが、改訂後の ISO 5725 からは削除された。一方 JIS では、電卓一つで簡単に計算が可能であることから現場で良く利用されてきた現実もあり、相変わらず Grubbs の方法を第一とし Dixon の方法を第二としている。

一方、最大値も最小値も両方疑わしい場合であると、ISO 5725 は Pearson - Stephens の方法[3]を推奨している（図 11）。この方法では、Grubbs の方法の T_1 や T_n のかわりに以下の式で計算される R/σ を用いる。

$$\frac{R}{\sigma} = \frac{x_n - x_1}{\sigma} \tag{41}$$

Grubbs の方法と同様に有意水準 α とデータ数 n に対応した棄却限界を与える数表が示されており（表 2）、表の値以上であれば最大値と最小値両方を棄却することになる。

ここまでの説明で、棄却検定は管理図の応用とよく似ていることに気がつかれた方もおられると思う。実際、棄却検定の手法は、過去のデータの蓄積にまさるものではない。したがって、過去のデータの蓄積（経験）により作製された管理図が存在する場合には、そちらによる判断を優先させるのも一つの方法といえよう。また Grubbs の方法では、データ数を無限大とした場合の有意水準に対する棄却限界の数表も示されており、管理図との併用には、こちらの方がよいかもしいない。

棄却検定は、「分散分析」と呼ばれる統計処理のもっとも基本となるものである。その前段階として、測定に付随する誤差をあらかじめ十分に検討しておかねばならない。これに対して、データに対する許容誤差という概念とその標準的な手順については、JIS Z 8402 に示されている。

さて今回は、2つの平均値を比べて差が有為かどうかを見る、2組のデータのばらつきを見て同じかどうかを判断する、あるいはいったい測定を何回繰り返せばよいか、などのより実践的な話をしてみたい。

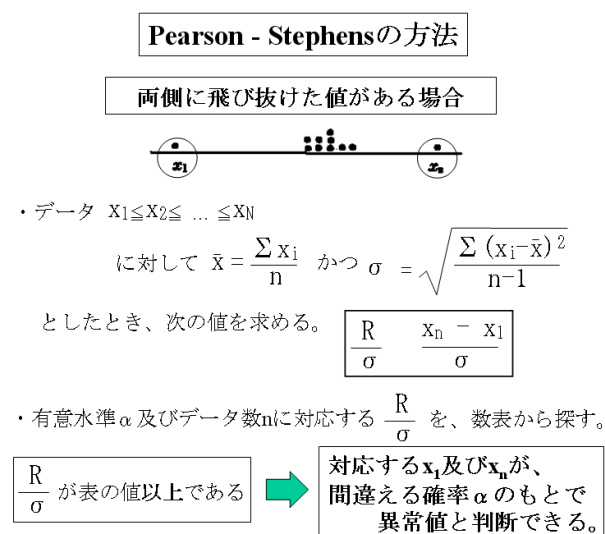


図 11 Pearson-Stephens の方法

参考文献

[1] F. E. Grubbs, G. Beck, Technometrics **14**, 847 (1972).
 [2] W. J. Dixon, Biometrics **9**, 74 (1953).
 [3] E. S. Pearson, M.A. Stephens, Biometrika **51**, 484 (1964).

表 2 Pearson-Stephens の棄却検定表

| データ数 n | 上側確率(α) | | | |
|-----------|------------------|-------|-------|-------|
| | 0.05 | 0.025 | 0.01 | 0.005 |
| 3 | 1.999 | 2.000 | 2.000 | 2.000 |
| 4 | 2.429 | 2.439 | 2.445 | 2.447 |
| 5 | 2.753 | 2.782 | 2.803 | 2.813 |
| 6 | 3.012 | 3.056 | 3.095 | 3.115 |
| 7 | 3.222 | 3.282 | 3.338 | 3.369 |
| 8 | 3.399 | 3.471 | 3.543 | 3.585 |
| 9 | 3.552 | 3.634 | 3.720 | 3.772 |
| 10 | 3.685 | 3.777 | 3.875 | 3.935 |
| 11 | 3.80 | 3.903 | 4.012 | 4.079 |
| 12 | 3.91 | 4.02 | 4.134 | 4.208 |
| 13 | 4.00 | 4.12 | 4.244 | 4.325 |
| 14 | 4.09 | 4.21 | 4.34 | 4.431 |
| 15 | 4.17 | 4.29 | 4.44 | 4.53 |
| 16 | 4.24 | 4.37 | 4.52 | 4.62 |
| 17 | 4.31 | 4.44 | 4.60 | 4.70 |
| 18 | 4.37 | 4.51 | 4.67 | 4.78 |
| 19 | 4.43 | 4.57 | 4.74 | 4.85 |
| 20 | 4.49 | 4.63 | 4.80 | 4.91 |
| 25 | 4.71 | 4.87 | 5.06 | 5.19 |
| 30 | 4.89 | 5.06 | 5.26 | 5.40 |
| 35 | 5.04 | 5.21 | 5.42 | 5.57 |
| 40 | 5.16 | 5.34 | 5.56 | 5.71 |
| 45 | 5.26 | 5.45 | 5.67 | 5.83 |
| 50 | 5.35 | 5.54 | 5.77 | 5.93 |
| 55 | 5.43 | 5.63 | 5.86 | 6.02 |
| 60 | 5.51 | 5.70 | 5.94 | 6.10 |
| 65 | 5.57 | 5.77 | 6.01 | 6.17 |
| 70 | 5.63 | 5.83 | 6.07 | 6.24 |
| 75 | 5.68 | 5.88 | 6.13 | 6.30 |
| 80 | 5.73 | 5.93 | 6.18 | 6.35 |
| 85 | 5.78 | 5.98 | 6.23 | 6.40 |
| 90 | 5.82 | 6.03 | 6.27 | 6.45 |
| 95 | 5.86 | 6.07 | 6.32 | 6.49 |
| 100 | 5.90 | 6.11 | 6.36 | 6.53 |
| 150 | 6.18 | 6.39 | 6.64 | 6.82 |
| 200 | 6.39 | 6.60 | 6.84 | 7.01 |
| 500 | 6.94 | 7.15 | 7.42 | 7.60 |
| 1000 | 7.33 | 7.54 | 7.80 | 7.99 |