実験データを正しく扱うための実験・分析データの統計解析入門　まとめ

甲南大理工 山本雅博

September 30, 2014

1　初めに

測定には誤差が必ずともなう。誤差を減らすには測定法、その方法の限界、装置等を習熟していわゆるバイアスのかかった測定にならないように細心の注意を払わなければならない。正確な測定が可能になっても、ある程度の測定の不確かさ（誤差）は避けられない。その不確かさが測定毎にランダムであると仮定すれば、データを統計処理することによって、不確かさを評価し、測定回数を増やすことで不確かさを小さくすることができる。

1 回しか測定できない超新星爆発の観測や特別の場合を除いて，常に多数回（分野にもよるが，統計処理可能な３－５回以上）の実験を行いその再現性を確かめることが重要である。また，その多数回の実験から統計・誤差解析をおこなってどの程度の信頼性があるのかを明確にする必要がある。[1]誤差解析は、以上の理由から科学・技術のすべての分野で欠くことのできないものである。[2]

2　教科書・参考書

[0]「実験データを正しく扱うために」化学同人編集部編, 化学同人, 2007.　（通称「緑本」 現在 8 刷）山本が所属している学科では，この本を教科書にして３年生の前期に 15 回の講義を行っている。`http://www.chem.konan-u.ac.jp/applphys/web_material/data_handle/` にこの本に関する著者らのページあり

本講義でカバーできない点は，以下の参考書で補って頂きたい。以下の本を参考文献として推薦する。（適当な参考文献はそう多くない。）

[1] Daniel C. Harris, Quantitative Chemical Analysis, Chap.3-5.　6th Ed. 2003.

[2] J. R. Taylor (林、馬場　訳）計測における誤差解析入門、東京化学同人, 2000. （An Introduction to Error Analysis, 2nd Ed. 1997 Univ. Sci. Books)（統計分析的にはあいまいな部分も多いがわかりやすい。）

[3] 小寺平治　ゼロから学ぶ統計解析　講談社 2002 (確率・統計の勉強は数式の証明を追いかけることで精一杯になることが多い。数学が専門の筆者がその点を反省して書いた本。数学的な証明がないぶん，ざっと概念をとらえる本として非常によい。）

[4] 小針アキ宏　確率・統計入門　岩波書店 1973（数理統計の本といえばこの本。数学的な証明はこの本を参照頂きたい。広中平祐氏による序文もおもしろい。葡萄パンと Poisson 分布の話が時代を感じさせるとも。）

[5] 鳥居泰彦　はじめての統計学　日本経済新聞　1994（ワークブック形式になっているのでわかりやすい。経済学部の教科書である。）

[6] P. R. Bevington and D. K. Robinson, Data reduction and error analysis for the physical sciences, WCB/McGraw-Hill, 2nd Ed. 1992. [7] N. C. Barford （酒井　訳）実験精度と誤差、丸善、1997.

3　正確さと精密さ：accuracy　vs　precision

・accuracy（正確さ、正確度）: Measure of how close the result of an experiemt comes to the “true value”. 真の値に如何に近いかの尺度。

・precision（精密さ、精度）: Measure of how carefully the result is determined without reference to any “true” value.　一連の実験結果がどの程度一致しているかの尺度。真の値との関係は示してない。

・系統誤差 (systematic errors) : 再現性有り。一方向。繰り返し測定を行っても原因は明らかにはならないことが多い。

・ランダム誤差（random errors) : 制御できない種類の不規則誤差。統計的手法を用いれば、ランダム誤差については信頼性のある評価を行うことが可能である。

4　表記、有効数字

ある量 x に関する測定値は、最良推定値 (best estimate) $x_{\rm best}$ と測定にともなう不確かさまたは誤差 (error) $\delta x (\geq 0)$ で次のように表される。

$$x(\text{measured}) = x_{\rm best} \pm \delta x \tag{1}$$

[1]結果を公表した後に，他の研究者が追試・再現できない場合，公表した研究者はその理由を説明する義務がある。実験ノート・実験生データ・統計処理した結果・表・グラフ（コンピュータファイルの場合，ファイル名・フォルダーの整理が必要）が，確実に存在しないと検証・反論できない。これまで「捏造（実験を実際に行っていなかったり，データを改ざんしたりすること）」となって報道されている場合，これら自身が存在しないか，整理が非常に悪いことが多い。

さらに，実験が間違いであった場合は，結果を取り下げる必要がある。結果を取り下げた場合，その研究者（and/or その共同研究者）の信頼性は失われる。極端な場合，捏造（実験を実際に行っていなかったり，データを改ざんしたりすること）が発覚した場合研究者その業界で生きていくことは基本的にできない。

[2]統計処理の基礎をなす数理統計学をカリキュラムにいれてない学科がほとんどのようである。

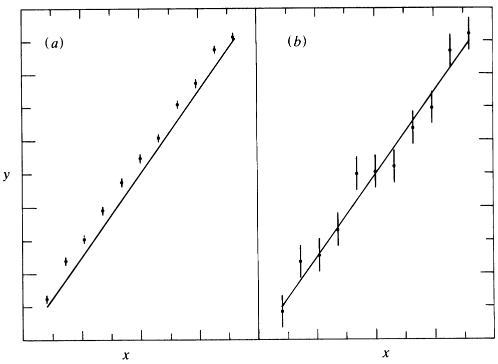


Figure 1: 実線は真の値。(a)：precise but inaccurate data. (b) accurate but imprecise data. 参考文献 [2] Fig.1.1(p.2) より転載。

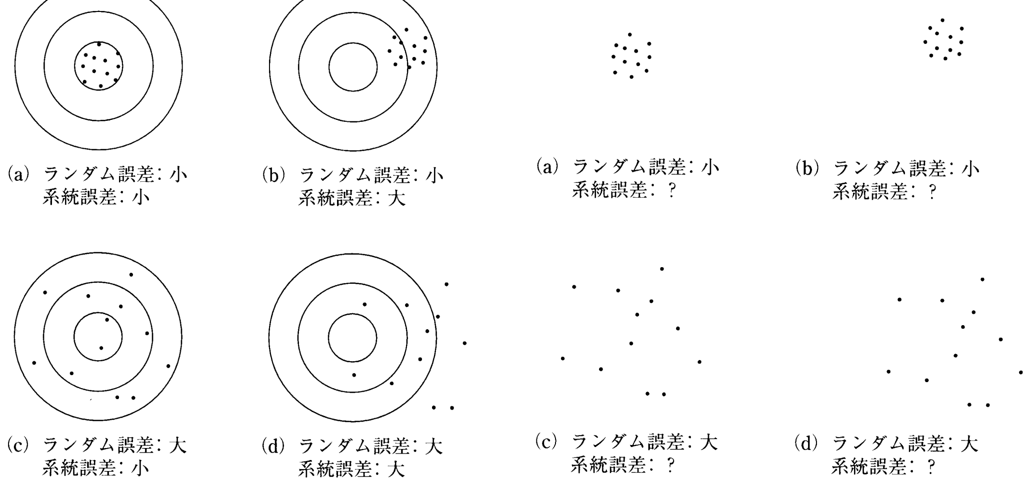


Figure 2: 左図：真の値がわかっている場合。Accuracy vs Precision も明確に定義できる。真の値が未知の一般的な測定（右図）においては、系統誤差は明らかでない。参考文献 [1]p.101-102 図４・1、図４・2 より転載。

相対誤差は

$$\text{relative uncertainty} = \frac{\delta x}{|x_{\text{best}}|} \tag{2}$$

で定義される。

計算で求められた誤差は有効数字１桁に丸め､最良推定値は、最終桁が誤差の最終桁と同じ位置になるように表す。

例：(1) 距離の測定で最良推定値（平均値）が 105.4 m で、計算された誤差が 0.2345 m ならば 105.4 ± 0.2345m ではなく，105.4 ± 0.2m と書くのが正しい。

(2) 同じ理由で， $105.4236 \pm 0.02\text{m} \Rightarrow 105.42 \pm 0.02\text{m}, 105.4236 \pm 2\text{m} \Rightarrow 105 \pm 2\text{m}$ である.

後述するように、有効数字・有効桁は誤差をいかに求めるかで決定され、 測定値の代数的な操作（加減乗除算，その他の演算）による有効数字・有効桁の変化は誤差伝搬の式から評価できる。

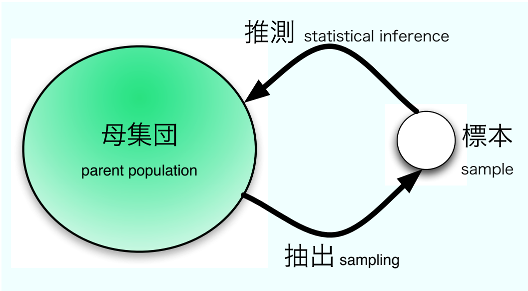
有効数字・有効桁は，小数点を移動させて，(符号)（x.xxx)×10^y の形をつかうとわかりやすい。すなわち，同じ 1234500 という数においても，有効桁５桁， ６桁， ７桁ではそれぞれ，$1.2345 \times 10^5, 1.23450 \times 10^5, 1.234500 \times 10^7$ と表記される。また，-0.00987600 の場合，有効桁４桁，５桁，６桁ではそれぞれ，$-9.876 \times 10^{-3}, -9.8760 \times 10^{-3}, -9.87600 \times 10^{-3}$ と表記される。

例：有効桁の概念に強く縛られると，２つの非常に近い測定値を引き算するとき混乱する。$(1.2345 \pm 0.0002) - (1.2343 \pm 0.0002) = 3 \times 10^{-4} \pm 3 \times 10^{-4}$ となり，有効桁は５桁から１桁に桁落ちする。（誤差が $\pm 3 \times 10^{-4}$ になる理由は後述する。）あくまで，有効桁は誤差の見積もりから判断すべきものである。有効桁５桁に固執すると，$3.0000 \times 10^{-4} \pm 3 \times 10^{-4}$ となり，誤差の桁より小さな数字が有効となり，明らかに矛盾する。

5　母集団，標本，抽出，推測

国勢調査や国産牛肉の全頭 BSE 検査のように統計をとるため全ての標本 (sample) を調べることが可能であればその集団の統計量は正確に求めることができる。通常は，その全体すなわち母集団 (parent population) から，無作為に標本を抽出（sampling）して，抽出された標本から母集団の統計量を推測することがなされている。

実験も同じである。系統的な誤差の小さい正確な測定が，１００回も１０００回も（異なる試料で）測定することが可能であるなら，得られた値は正確であろう。問題は，その測定を（できるなら，なるべく少ない回数で）何回行い，どのような統計処理を行えば，測定が如何に正確であるか，またそれを如何に他人に信用してもらうかである。



6 平均（標本平均）、分散、標準偏差、平均の平均・分散

ある量 x を同じ装置で同じ手順で N 回測定し、$x_1, x_2, , , , x_N$ の N 個の値を得たときその 平均値 (average or mean)[標本平均]$\bar{x}$ は，(言い換えるなら，母集団より N 個の標本を抽出したとき)

$$\bar{x} = \frac{x_1 + x_2 + ... + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

標本分散 (sample variance)s^2 は，以下の様に定義される。[3]

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{4}$$

また， 標準偏差 s(SD:standard deviation) は分散の平方根で定義する。

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{5}$$

例：(悪名高き) 偏差値の定義は，(得点ー平均点) /標準偏差 $\times 10 + 50$ で定義する。

ここで重要なことは，標本平均 $\bar{x}$ および標本分散 s^2 は，母集団の平均である 母平均 μ(population mean) (ミューとよむ) および 母分散 σ(popolation variance) （シグマとよむ）とは一致しないことである。($N \to \infty$ で一致する。)

N 個の標本値の平均をとることを 1 回の操作としてその操作を何回も行うとする。そうすると，$\bar{x}$ はひとつの分布 $f(\bar{x})$ をもつことになる。この分布 f から， 平均の平均 , 平均の分散 が定義できる。

例えば，5 回の測定を 230 人の学生が行うとき，各人の求めた平均値は分布をもつ。また，ある測定を 5 回繰り返し平均値を求め，その測定を別の試料でさらに 3 回繰り返した場合も平均値は分布をもつ。いずれにしても測定は独立（前の測定が後の測定に影響を与えない。）であるとすれば，以下の中心極限定理がなりたつ。

7 「平均の平均」に対する中心極限定理

任意の形の分布をもつ母集団が，平均 μ, 分散 σ^2 をもつとする。**標本サイズ N が十分大きいとき**，標本平均 $\bar{x}(= \sum_i^N x_i/N)$ は，近似的に正規分布 $G_{\mu,\sigma/\sqrt{N}}(\bar{x})$ に従う。

8 正規分布（ガウス分布）

ランダムな誤差要因が有る場合の極限分布 ($N \to \infty$) は正規分布（またはガウス分布）$G_{\mu,\sigma}(x)$ に従うと考えてよい。[4]

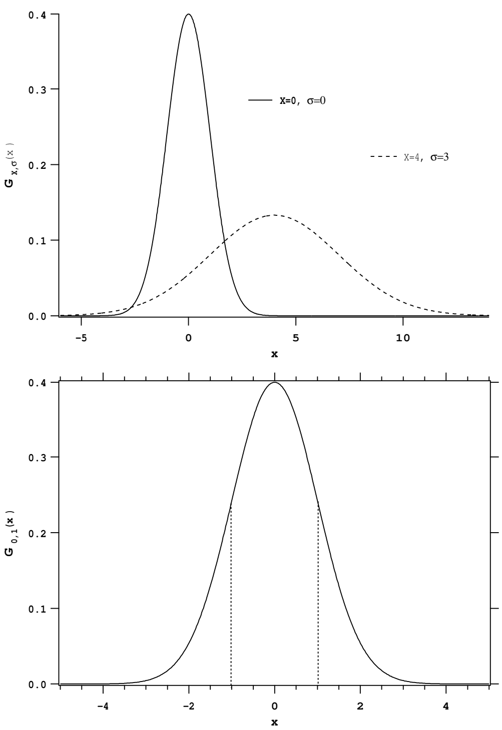
$$G_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \tag{6}$$

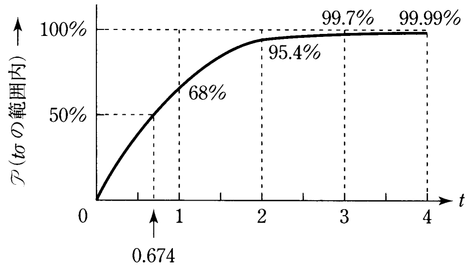
Fig. 3 に正規分布 $G_{\mu,\sigma}(x)$ を図示する。例：ちなみに偏差値 50 は平均，60 は top 16%, 65 は top6.5%, 70 は top2.3%, 80 は top0.15%である。年間 70 万人大学に入学するなら，偏差値 65 以上の人は 45500 人いることになる。[5]

[3] 分母の N を $N - 1$ で定義する場合もあるが，後に議論する。

[4] 表と裏のでる確率がそれぞれ 1/2 であるコインを投げて、表が出れば右に ϵ、裏がでれば $-\epsilon$ つまり左に ϵ 進むとする。n 回投げて r 回裏がでて $n - r$ 回表が出れば、$(n - 2r)\epsilon$ のところに居るわけで、その確率は ${}_n\mathrm{C}_r(1/2)^r(1/2)^{n-r}$ である。n が十分大きいと $(n - 2r)\epsilon$ にいる確率はガウス分布になることが数学的に示される。(**中心極限定理**)

[5] 偏差値というのは閉じられた集団での相対的な位置を示すにすぎない。集団全体の学力がさがっている場合，それに同調して学力がさがっても同じ偏差値を保つことができる。





t	0	0.25	0.5	0.75	1.0	1.25	1.5	1.75	2.0	2.5	3.0	3.5	4.0
$\mathcal{P}$[%]	0	20	38	55	68	79	87	92	95.4	98.8	99.7	99.95	99.99

Figure 3: 左上図：２つのガウス分布。左下図：$\mu \pm \sigma$ の間の部分は、平均値 μ から標準偏差 1 個分の範囲内に x が入る確率。右図：平均値 μ から標準偏差 t 個分内の範囲に x が入る確率。この関数は誤差関数 erf(t) と呼ばれる。右図は参考文献 [1] 図 5・13（p.145) より転載。

9　独立な多変数分布から求めた「平均の平均」と「平均の分散」

いま，(x, y) という 2 変数の確率分布を $p(x, y)$ を考える。x と y の関数 $h(x, y)$ のこの分布における期待値 $\langle h \rangle$ は,

$$\langle h \rangle = \int\int dxdyh(x,y)p(x,y) \tag{7}$$

x と y が独立であれば,

$$p(x, y) = f(x)g(y) \tag{8}$$

ただし,

$$f(x) = \int dyp(x,y) \tag{9}$$

$$g(y) = \int dxp(x,y) \tag{10}$$

である。

いま,$h(x, y) = x + y$ で，x と y が独立であるとしよう。

$$\begin{aligned}\langle x+y \rangle &= \int\int dxdy(x+y)f(x)g(y) = \int dxxf(x)\underbrace{\int dyg(y)}_{=1} + \int dyyg(y)\underbrace{\int dxf(x)}_{=1} \\ &= \langle x \rangle + \langle y \rangle \end{aligned} \tag{11}$$

同様にして, 以下の式が得られる。

$$\frac{1}{N}\langle x_1 + x_2 + x_3 + ... + x_N \rangle = \frac{\langle x_1 \rangle + \langle x_2 \rangle + \langle x_3 \rangle + ... + \langle x_N \rangle}{N} \tag{12}$$

$\langle x_i \rangle = \mu$(母集団の平均) であるとすれば，上式の左辺に相当する「平均の平均」$\langle \bar{x} \rangle$ は ($\langle x \rangle$ と $\bar{x}$ の意味は違うことに注意。)

$$\langle \bar{x} \rangle \equiv \langle \frac{x_1 + x_2 + x_3 + ... + x_N}{N} \rangle = \frac{N\mu}{N} = \mu \tag{13}$$

となる。

次に，$h(x,y) = [(x+y) - \langle x+y\rangle]^2$ を考える。x と y は独立であるとしよう。

$$\begin{aligned}
\langle [(x+y) - \langle x+y\rangle]^2\rangle &= \int\int dxdy[(x+y) - \langle x+y\rangle]^2 f(x)g(y) \\
&= \int dxx^2 f(x) + \int dyy^2 g(y) + 2\int dxxf(x)\int dyyg(y) - 2\langle x+y\rangle^2 + \langle x+y\rangle^2 \\
&= \langle x^2\rangle + \langle y^2\rangle + 2\langle x\rangle\langle y\rangle - \langle x\rangle^2 - \langle y\rangle^2 - 2\langle x\rangle\langle y\rangle = \langle x^2\rangle - \langle x\rangle^2 + \langle y^2\rangle - \langle y\rangle^2 \qquad (14)
\end{aligned}$$

従って，以下の式が得られる。

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 \qquad (15)$$

$$\sigma_x^2 = \int dx(x - \langle x\rangle)^2 f(x) = \langle x^2\rangle - \langle x\rangle^2, \quad \sigma_y^2 = \int dy(y - \langle y\rangle)^2 g(y) = \langle y^2\rangle - \langle y\rangle^2 \qquad (16)$$

同様にして

$$\langle [(\frac{x_1 + x_2 + x_3 + ... + x_N}{N}) - \langle \frac{x_1 + x_2 + x_3 + ... + x_N}{N}\rangle]^2\rangle = \frac{\langle x_1^2\rangle - \langle x_1\rangle^2 + ... + \langle x_N^2\rangle - \langle x_N\rangle^2}{N^2} \qquad (17)$$

$\langle x_i^2\rangle - \langle x_i\rangle^2 = \sigma^2$(母集団の分散）であるとすれば, 以下の重要な式が得られる。

$$\sigma_{\bar{x}}^2 = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N} \qquad (18)$$

10　誤差伝播（でんぱ）

測定値 $x,,,,z$ の測定誤差が $\delta x,,,,\delta z$ であるとき、測定値を用いて関数 $q(x,,,,,z)$ の値を計算すると $x,,,,z$ の誤差が互いに独立かつランダムであるとすれば q の誤差は、

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + ... + \left(\frac{\partial q}{\partial z}\delta z\right)^2} \qquad (19)$$

で与えられる。また、互いに独立であるかないかに関わらず以下のような通常の和より大きくなることはない。

$$\delta q \le \left|\frac{\partial q}{\partial x}\right|\delta x + ... + \left|\frac{\partial q}{\partial z}\right|\delta z \qquad (20)$$

証明は、

$$q(x;\delta x, y + \delta y) \simeq q(x,y) + \left(\frac{\partial q}{\partial x}\right)\delta x + \left(\frac{\partial q}{\partial y}\right)\delta y + ... \qquad (21)$$

で、右辺第２項は中心が 0 で幅が $\left(\frac{\partial q}{\partial x}\right)\delta x$ のガウス分布している。第 3 項もおなじ。x,y がガウス分布している場合、その和 $x+y$ もガウス分布しその標準偏差は $\sqrt{\sigma_x^2 + \sigma_y^2}$ になることが知られているので、上の誤差伝播の式が成り立つ。

問題：(1)　ライスを 500 ± 20 g, カレールーを 100 ± 10 g 盛りつけてくれる食堂では，カレーライスとして何 g ± 何 g 盛りつけてくれるといえるのか?[参考文献 4, p94] [6]**解答：**$q = x+y = 600, \partial q/\partial x = 1, \partial q/\partial y = 1, \delta q = \sqrt{(\delta x)^2 + (\delta y)^2} = \sqrt{20^2 + 10^2} = 20$ (計算では 22.4 となるが 誤差は 1 けたにまとめる。), Ans. $600 \pm 20g$

(2) さらにカレーに福神漬けを，1.0 ± 0.5 g つけてくれるとしたら，総計何 g ± 何 g 盛りつけてくれるといえるのか？ **解答**：600 ± 20 g（1.0 ± 0.5 g は誤差 20g 以下）

(3) $(7\pm2)\times k, k =$Constant :**解答** $q = kx, \partial q/\partial x = k, \delta q = k\delta x$, Ans. $7k \pm 2k$.

問題 (4) (5±1)×(8±2), (5) (10±1)/(20±2)　を $x \pm \delta x$ の形で表せ。

11　母集団の平均の推定：母集団の分散が既知の場合

母集団がガウス分布をしていてその分散 σ^2 が既知の場合 (例：精密電子天秤での重量測定の場合)，サイズ N の標本（N 回の実験）から標本平均 $\bar{x}$ を求める。母集団の平均は，信頼度 (confidence level)68.3%, 90%, 95%の信頼区間 (confidence interval) で, すなわち母集団の平均がこの範囲に入っているのが 68.3%, 90%, 95%確からしいのは,

$$\bar{x} - z\frac{\sigma}{N} \le \mu \le \bar{x} + z\frac{\sigma}{N}, \quad z = 1(68.3\%),\ z = 1.645(90\%),\ z = 1.960(95\%) \qquad (22)$$

[6]和の分布と混合の分布は異なる。参考文献 [5, 118-119] にある例はおもしろい。背の高い北欧人と背の低いアフリカの某族の身長の和の分布を考える。双こぶの分布になると混同しやすいが、それは混合したグループの分布である。和の分布は北欧人の頭に某族の人をのせたトーテムポールの高さの分布である。

母集団がガウス分布していなくても，標本サイズ N が十分大きい場合は，中心極限定理により母集団の平均は上の式を満たす。以上より、測定値の平均および誤差（エラーバー）

$$x(\text{measured}) = \bar{x} \pm z\frac{\sigma}{\sqrt{N}}, \quad z = 1(68.3\%),\ z = 1.645(90\%),\ z = 1.960(95\%) \tag{23}$$

が得られる。ここで信頼度に関係する z の値は Fig.3 の右図より得られる。信頼度としては，統計解析の世界では 95%または 90%をとるのが慣例となっているが，場合によっては 68.3%で構わない。(信頼度を明示すればよい。）**N 回測定することにより平均値の不確かさは $1/\sqrt{N}$ で減少する。**実験データは、平均値とエラーバー（信頼限界，confidence interval) で示されるがエラーバーの大きさは $z\sigma_x/\sqrt{N}$ となる。(Fig.4) 従って，信頼度を明示してエラーバーを記述する必要がある。教科書の Fig.4-5 を参照のこと。

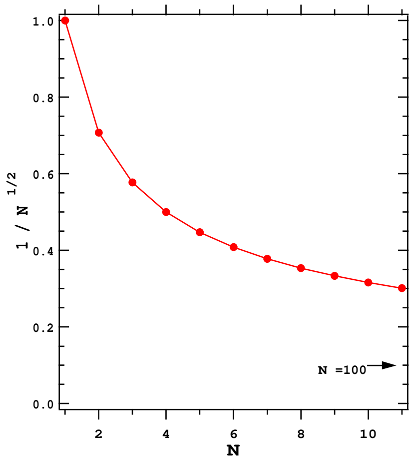


Figure 4: N vs $1/\sqrt{N}$

ここで，注意しなくてはならいないのは，σ^2 は，標本分散 $s^2 = (1/N)\sum_i^N (x_i - \bar{x})^2$ ではなく，あくまで既知の母集団の分散である ということである。以下で，母集団の分散が未知である場合について考える。

12　母集団の分散が未知の場合の平均の推定

実際の実験においては，母集団の分散 σ^2 を事前に知ることはまずない。従って，以下に述べる方法により，平均値の信頼区間 (言い換えるならエラーバー）を求めるのが通常である。まず，標本分散 $s^2 = (1/N)\sum_i^N (x_i - \bar{x})^2$ の平均をもとめよう。この場合も，N 個の標本から s^2 を計算することを 1 回の操作と考えて，その操作を多数回繰り返せば s^2 に関する分布ができるその平均・分散が定義できる。

今，s^2 を以下のように変形して考える。

$$s^2 = \frac{1}{N}\sum_i^N (x_i - \bar{x})^2 = \frac{1}{N}\sum_i^N (x_i - \mu + \mu - \bar{x})^2 = \frac{1}{N}\sum_i^N [(x_i - \mu)^2 + 2(\mu - \bar{x})(x_i - \mu)] + (\mu - \bar{x})^2 = \frac{1}{N}\sum_i^N (x_i - \mu)^2 - (\mu - \bar{x})^2 \tag{24}$$

ここで，$\bar{x}, s^2$ は標本平均，標本分散である。また，母集団は，平均 μ, 分散 σ^2 をもつとしよう。平均 $\langle s^2 \rangle$ をとると，

$$\langle s^2 \rangle = \langle \frac{1}{N}\sum_i^N (x_i - \mu)^2 - (\mu - \bar{x})^2 \rangle = \frac{N\sigma^2}{N} - \frac{\sigma^2}{N} = \frac{N-1}{N}\sigma^2 \tag{25}$$

ここで，平均 $\bar{x}$ の分散 $= \sigma^2/N$ を第二項に対して使った。標本平均の平均は母集団の平均 μ と一致したが，標本分散の平均は母集団の分散 σ^2 とは一致しない。

標本分散 s^2 を変形して

$$u^2 = \frac{N}{N-1}s^2 = \frac{N}{N-1}\frac{1}{N}\sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N-1}\sum_{i=1}^N (x_i - \bar{x})^2 \tag{26}$$

を新たに定義する。u^2 を不偏分散 (unbiased variance) という。[7]

[7] N 個の標本から，$\sum_{i=1}^N (x_i - \bar{x})^2$ 求める際，$\bar{x} = (1/N)\sum_{i=1}^N x_i$ の条件が付くため，$(x_i - \bar{x})^2$ は N 個の独立な項の和ではなく，独立なのは（すわなち自由度）は $N-1$ 項である。

$G_{\mu,\sigma}$ のガウス分布をもつ母集団から得たサイズ N の標本の不偏分散を u^2 とすれば，$T = (\bar{x} - \mu)/(u/\sqrt{N})$ は，自由度 $N-1$ の Student の t 分布に従う。 自由度 m(m は 1 以上の整数）の Student[8]の t 分布 $f_m(t)$ は，

$$f_m(t) = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{m\pi}\Gamma(\frac{m}{2})}\left(1+\frac{t^2}{m}\right)^{-(m+1)/2} \tag{27}$$

と定義される。ここで，Γ はガンマ関数で，$\Gamma(1) = 1, \Gamma(n+1) = n!, \Gamma(1/2) = \sqrt{\pi}, \Gamma(n+1/2) = (2n)!/[2^{2n}n!]\sqrt{\pi}$ である。Fig.5 にその分布を示す。 Fig 中に示した $t_1(68.3\%)$ は，自由度 1 の分布 $f_1(t)$ において，２本の縦点線間での確率が 68.3%であることを示している。**自由度 m が小さいときは分布の幅は広がることに注意して欲しい。** m が無限大の時,$f_m(t)$ はガウス分布 $G_{0,1}(t)$ に一致する。

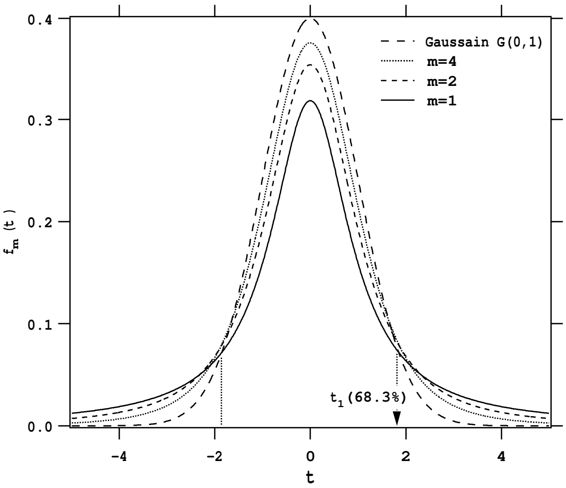


Figure 5: Student t 分布。自由度 m が小さいときは分布の幅は広がる。m が無限大の時,$f_m(t)$ はガウス分布 $G_{0,1}(t)$ に一致する。

この定理から以下の事が導き出せる。分散が未知の母集団からサイズ N の標本をとった時，標本平均を $\bar{x}$, 不偏分散を u^2 とすると，母集団の平均 μ の信頼度 z%での信頼区間は

$$\bar{x} - t_{N-1}(z\%)\frac{u}{\sqrt{N}} \le \mu \le \bar{x} + t_{N-1}(z\%)\frac{u}{\sqrt{N}} \tag{28}$$

で与えられる。

すなわち，**実験を N 回行って** $\{x_1, x_2, x_3, ..., x_N\}$ **のデータを得た時，その測定値は以下の様に評価される。**[9]

この式が最重要ポイント !!

$$x(\text{measured}) = \bar{x} \pm t_{N-1}(z\%)\frac{u}{\sqrt{N}} \tag{29}$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad u = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$t_{N-1}(z\%)$ は，Eq.(27) 式から求めることができるが,

例えば，実験を 5 回行った時の信頼度68.3%でのエラーバーは，$\pm 1.141u/\sqrt{5} = \pm 0.510u$ であり，実験を 3 回行った時の 95% 信頼度でのエラーバーは，$\pm 4.303u/\sqrt{3} = \pm 2.484u$ となる。 $t_{N-1}(z\%)/\sqrt{N}$ の N 依存性を Fig.6 に示す。Fig.4 と比較しよう。

[8]W. S. Gosset のペンネーム。Gosset の雇い主であるギネス社 (ビール会社）は，内容にかかわらずいかなる論文も発表することを雇用者に禁じたため（それ以前にギネス社の他の研究者が企業秘密を論文で公開してしまったため。), Gosset は Student というペンネームをつかって会社にばれないように公表した。“The probable error of a mean”, *Biometrika*, **1908** *6*, 1-25. (27) 式の証明は小針氏の本を参考にして頂きたいが，ガンマ関数，ベータ関数，χ^2(カイ二乗）分布，F 分布等の概念の後に t 分布が出てきて，数式を追いかけることで容易に挫折しやすいので省略した。(著者は某大学某学科の以前のカリキュラムで，2 回生の「数理統計学」および 3 回生の「技術者のための統計的品質管理入門」を履修したが，挫折組が多かったように思う。）どうしても気になる人以外は，(27) 式をとりあえず認めて頂きたい。

[9]当然ながら $N = 1$ では評価できない。

Table 1: Student の t 分布における $t_{N-1}(z\%)$ の表

自由度 (N-1)	標本数 (N)	$t_{N-1}(z\%)$, $z(\%)$:信頼度			
		68.3 %	90 %	95 %	99 %
1	2	1.837	6.314	12.706	63.657
2	3	1.321	2.920	4.303	9.925
3	4	1.197	2.353	3.182	5.841
4	5	1.141	2.132	2.776	4.604
5	6	1.110	2.015	2.571	4.032
6	7	1.090	1.943	2.447	3.707
7	8	1.077	1.895	2.365	3.500
8	9	1.066	1.860	2.306	3.355
9	10	1.059	1.833	2.262	3.250
10	11	1.052	1.812	2.228	3.169
15	16	1.034	1.753	2.131	2.947
20	21	1.026	1.725	2.086	2.845
25	26	1.020	1.708	2.060	2.787
30	31	1.017	1.697	2.042	2.750
40	41	1.013	1.684	2.021	2.704
60	61	1.008	1.671	2.000	2.660
120	121	1.004	1.658	1.980	2.617
∞	∞	1.000	1.645	1.960	2.576

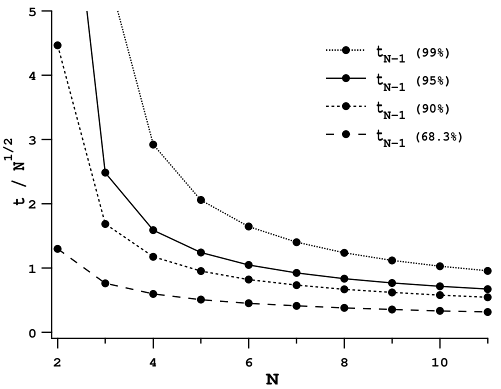


Figure 6: N vs $t_{N-1}(z\%)/\sqrt{N}$

以上では、分散が未知の場合の母集団の平均の信頼区間を求めたが、分散が未知の母集団の分散 σ^2 の信頼区間を χ^2(カイ二乗) 分布から求めることができる。詳細は紙面の都合で省略するが，標本数を N，標本分散を s^2 とすれば，信頼度 95%での信頼区間は，

$$\frac{Ns^2}{\chi^2_{N-1}(0.025)} \leq \sigma^2 \leq \frac{Ns^2}{\chi^2_{N-1}(0.975)}, \qquad s^2 = \frac{1}{N}\sum_i^N (x_i - \bar{x})^2 \tag{30}$$

である。自由度に依存した $\chi^2_{N-1}(0.025), \chi^2_{N-1}(0.975)$ は，参考書に表が記載されている。($\chi^2_4(0.025) = 11.14, \chi^2_4(0.975) = 0.484$ である。）興味のある人は参考文献を参照頂きたい。

13　仮説検定：母集団の平均値の検定

以上の議論は仮説の検定の基礎にもなっている。方法は，今，主張したい仮説があるとする。その仮説（対立仮説）を検定するのではなく，否定したい仮説（帰無仮説）が統計的にありえないことで棄却されることを利用する。従って，否定したい仮説が棄却される時は意味があるが，棄却されないときは「何もいえない」ことに注意したい。(二重否定は肯定ではない。「実験をやりたくないわけではない。」≠「実験をやりたい。」）仮説が偽であることを示すのはたやすいが，真であることを示すのは難しいという考え方である。

ここでは２つの例をあげよう。

1)　A 君は，食堂ではいつも大ライスを食べることにしているが，ある特定の人がもりつけてくれる時はいつも少ないとこぼしている。これを検証するために，その特定の人に５回大ライスを注文してその重さをはかったところ 352, 360, 352, 351, 357 g となった。大ライスは 360 g と表示されている。さて，この５回の測定で少ないといえるのだろうか？

この場合，上で述べた t 分布を用いて検定ができる。$T = (\bar{x} - \mu)/u/\sqrt{N}$ を求めて，それが 95%信頼区間の信頼区間外にあれば，平均が μ になるという否定したい仮説（帰無仮説）を棄却できる。$N = 5, \bar{x} = 354.(4), \mu = 360, u = [(352 - 354.(4))^2 + (360 - 354.(4))^2 + ..]/(5 - 1) = 3.9, T = (\bar{x} - \mu)/u/\sqrt{N} = -3.2$ となり，片側 5% で検定する（左側検定）ため，自由度 4 の 90%

信頼度区間-2.132 の値をつかうと $-3.2 < -2.132$ となり，帰無仮説は棄却される。すなわち，盛りつけは少ない言える。ただし，片側 0.5% で検定するすると，$-3.2 > -4.60$ となり，帰無仮説は棄却されず，なんとも（どちらとも）言えなくなる。

2) S 君は，[10] 空気から酸素を取り除いた後の窒素の量と化学的に発生させた窒素の量には Table3 に示すような差があることを見いだした。それは実験誤差に起因するのではなく，空気中に酸素・窒素より重い成分があるからであるという仮説を立てた。さて，もし本当であれば世紀の大発見ともなるこの仮説は，実験誤差に起因しないといえるのか？

Table 2: mass of constant volumes of gas (at constant temperature and pressure) isolated by romoving oxygen from air or generated by chemical decomposition from nitrogen compounds.[参考文献 1, p70]

From air / g	2.31017	2.30986	2.31010	2.31001	2.31024	2.31010	2.31028	
From chemical decomposition /g	2.30143	2.29890	2.29816	2.30182	2.29869	2.29940	2.29849	2.29889

今ガウス分布 $G_{\mu_A,\sigma_A}, G_{\mu_B,\sigma_B}$ をしている２つの母集団 A,B を考える。それぞれの母集団の分散 σ_A^2, σ_B^2 は未知であるが，本質的に等しいとする。それぞれの母集団からサイズ N_A, N_B 個の標本をサンプリングし，その標本平均，不偏分散をそれぞれ，$\bar{x}_A, \bar{x}_B, u_A^2, u_B^2$ としよう。この時，T_{AB} は，自由度 $(N_A + N_B - 2)$ の t 分布に従う。

$$T_{AB} = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\left(\frac{1}{N_A} + \frac{1}{N_B}\right) u_{AB}^2}}, \quad u_{AB}^2 = \frac{(N_A - 1)u_A^2 + (N_B - 1)u_B^2}{(N_A - 1) + (N_B - 1)} = \frac{\sum_i^{N_A}(x_{i,A} - \bar{x}_A)^2 + \sum_j^{N_B}(x_{j,B} - \bar{x}_B)^2}{N_A + N_B - 2} \tag{31}$$

このことを利用して，上の仮説の検証が可能になる。すなわち,「母集団 A,B の平均は等しい」ということを否定したい仮説（帰無仮説）とし，それを棄却できるか否かを T_{AB} を求め，95%の信頼性区間の外にあるか否かで決定する。$N_A = 7, \bar{x}_A = 2.31011, u_A = 0.00014(3), N_B = 8, \bar{x}_B = 2.29947, u_B = 0.00138$ であり，$u_{AB} = 0.00102, T_{AB} = (2.31011 - 2.29947)/0.00102/\sqrt{(1/7) + (1/8)} = 20.2$ となる。この値は自由度 13(=7+8-2) の値 2.160 よりも遙かに大きく，**差がないという仮説は棄却される**。（信頼度を 99.99% にしても 5.5 であり，帰無仮説は棄却される。）

14　観測値の棄却基準: Ｑテスト

データの棄却の基準には様々な方法があるがここでは、Ｑテスト [R. B. Dean and W. J. Dixon, Anal. Chem., 23, 636 (1951).] の具体的方法について述べる。その理論的裏付けはここでは述べない。

１）測定値を数値の順に並べ替える。

２）X =｜疑わしい測定値 (max or min) - その隣の数値｜ / （最大値－最小値）

３）X が統計的に計算されている数値（Q 値）と比較して等しいか大きい場合棄却する。

90%の信頼限界でのＱ値を以下の表に示す。

Table 3: 90%信頼限界における棄却係数（Ｑ値）

測定回数	3	4	5	6	7	8	9	10
Ｑ値	0.94	0.76	0.64	0.56	0.51	0.47	0.44	0.41

例として、10 回の測定により測定値、-3.4, -2.1, -1.7, -1.4, -1.4, -1.3, -1.1, -0.6, 8.5, -1.6 が得られたとしよう。10 回の測定であるので、上の表から Q 値は 0.41 である。この 10 個のデータ中、棄却の対象となりそうなのは最大値の 8.5 と最小値-3.5 である。それぞれの X 値は $|-3.4 - (-2.1)|/|8.5 - (-3.4)| = 0.11, |8.5 - (-0.6)|/|8.5 - (-3.4)| = 0.76$ となり、測定値-3.4 に対しては Q 値以下なので棄却しないが、測定値 8.5 に対しては Q 値以上なので棄却することになる。測定値の平均値は、従って、8.5 を除いた他の９つの値の平均-1.6 となる。

15　最後に

その他、母集団の分散に関する検定、F 分布，適合度の検定 (χ^2 検定)，最小２乗法 (回帰分析)、相関分析等については，紙面の都合で省略した。このプリントの知識で十分理解可能であるので，興味のある人は参考文献の本を参照頂ければ幸いである。

[10] John William Strutt、レイリー卿の通称で知られる。光の散乱の研究から空が青くなる理由を示す（レイリー散乱）、地震の表面波（レイリー波）の発見、ラムゼーとの共同研究によるアルゴンの発見、熱放射を古典的に扱ったレイリー・ジーンズの法則の導出などを行った。このほかにも流体力学（レイリー数）や毛細管現象の研究など、古典物理学の広範な分野に業績がある。「気体の密度に関する研究、およびこの研究により成されたアルゴンの発見」により、1904 年の ノーベル物理学賞を受賞した。http://ja.wikipedia.org/wiki/ジョン・ウィリアム・ストラット　より一部抜粋。